# Lecture 8: Multivariate Time Series Models

Nan-Jung Hsu

May 21, 2020

# Textbooks

- Time Series Analysis and its Applications (SS), chapters 5-6
- Multivariate Time Series Analysis (Tsay), chapters 1-2

# Multivariate Time Series: Overview

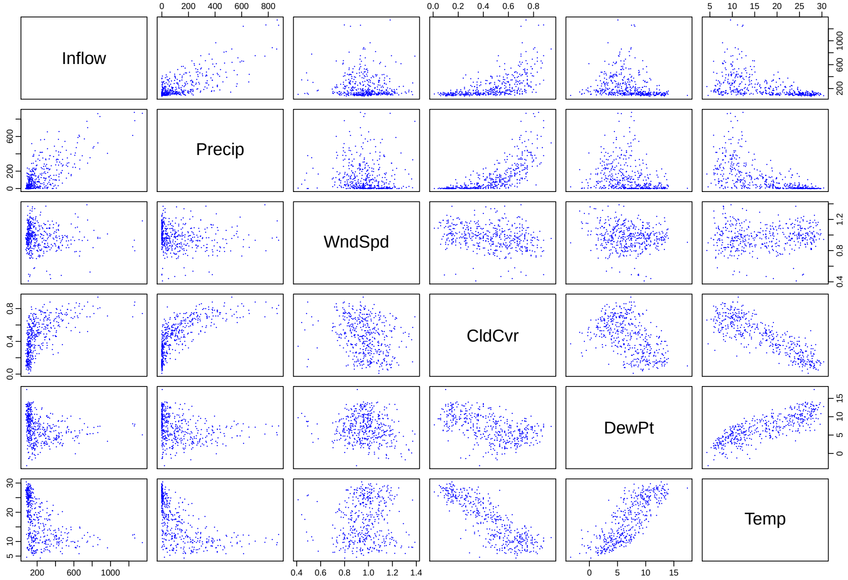
Why need jointly modeling more than one series?

- want to know the relationship between series
  - Granger causality
  - instantaneous causality
- enable to provide timely and more efficient forecasts (borrowing information across correlated series)
  - regression concepts
  - leading indicator (transfer function model)
  - co-integration (stable relationship among series)

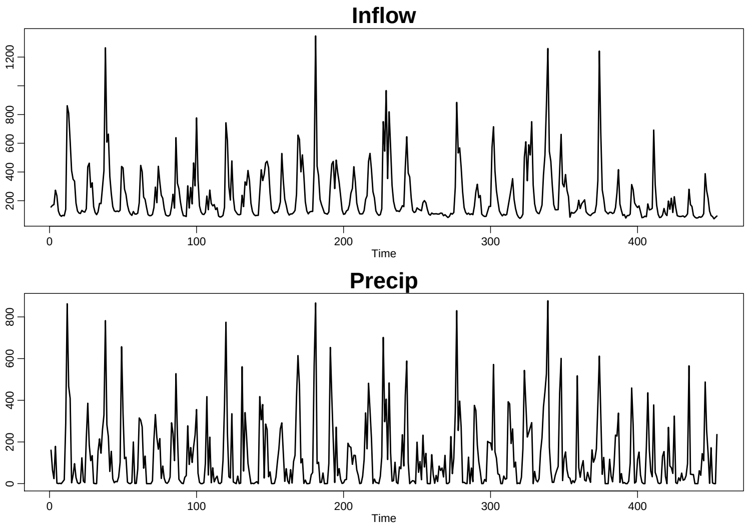
Popular multiple time series models:

- Vector AR model
- transfer function model
- factor model (state-space model)

# Example 1: Lake Shasta Inflow Data (monthly data $T = 454$)

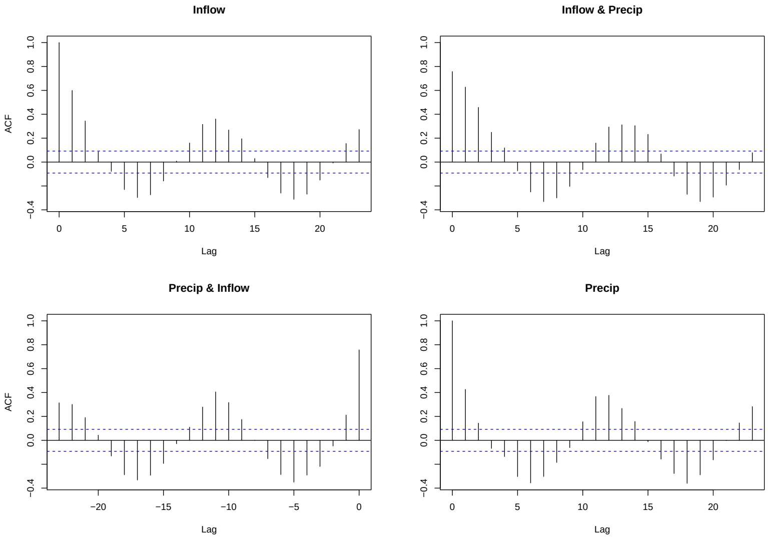


# Example 1: Lake Shasta Inflow Data (cont.)

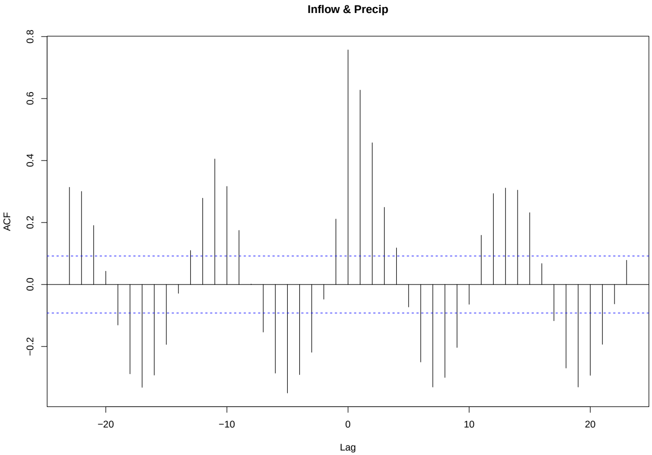



Clear co-movement in 2 series with fairly regular periodic patterns.

# Example 1: Lake Shasta Inflow Data (cont.)

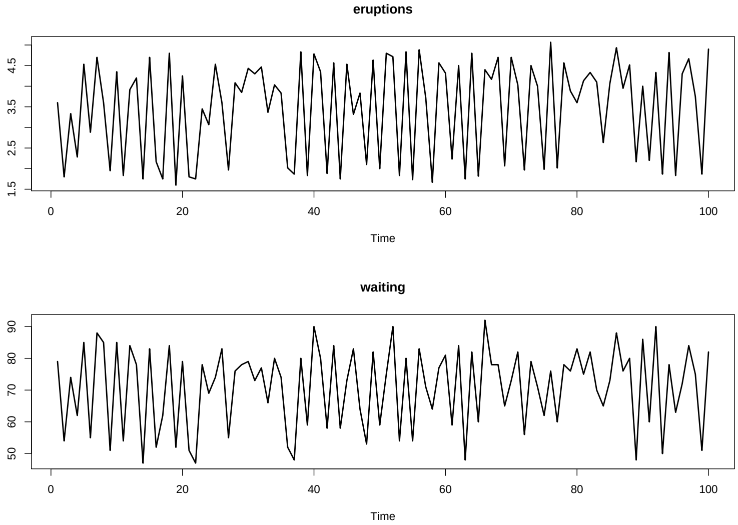


# Example 1: Lake Shasta Inflow Data (cont.)



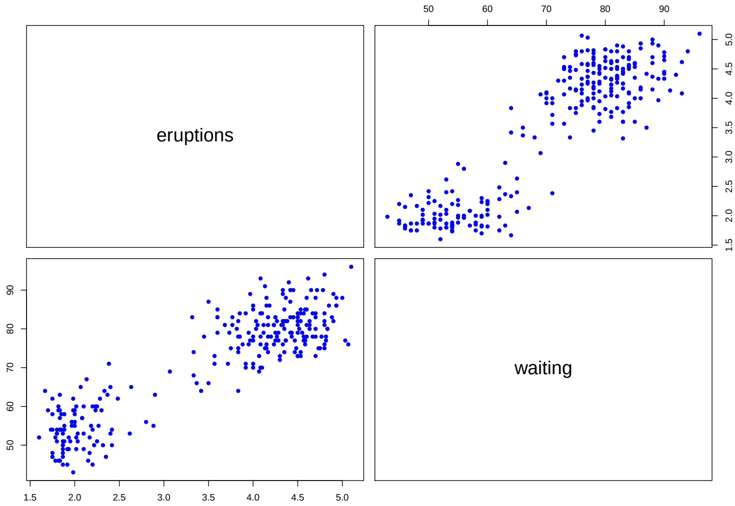

Sample CCF indicates: Dynamic relationship between 2 series is not symmetric!

# Example 2: Old Faithful Data (duration time in minutes)



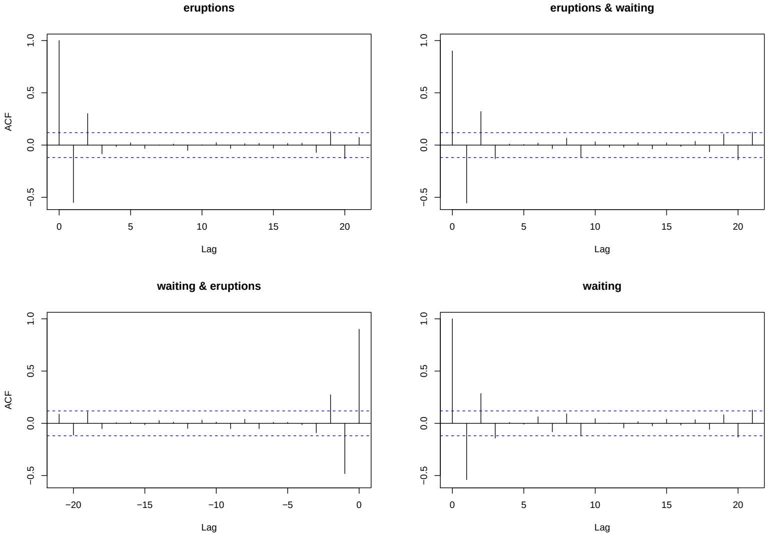

Clear co-movement in 2 time series with negative lag-1 correlation.

# Example 2: Old Faithful Data (cont.)



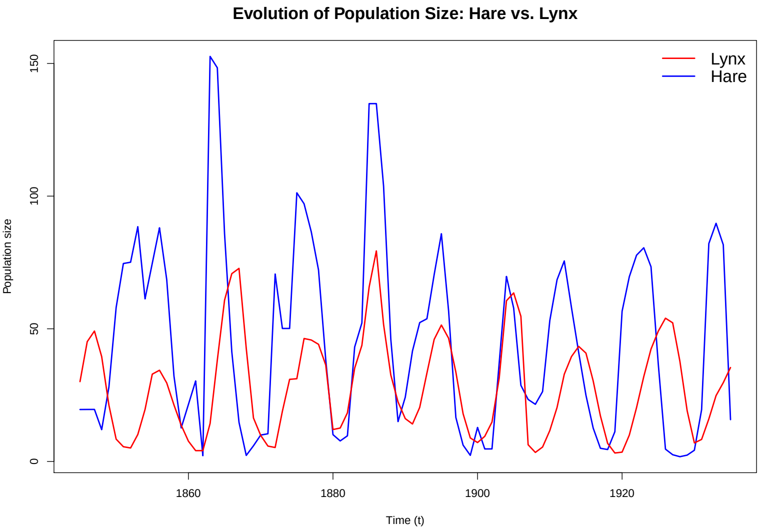

Eruption time and waiting time are concordant but distributed in 2 well-separated groups.

# Example 2: Old Faithful Data (cont.)

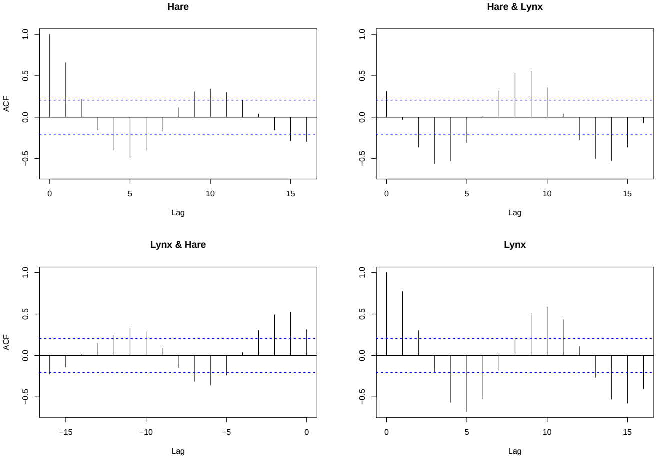



Sample CCF indicates: Dynamic relationship between 2 series is fairly symmetric!

# Example 3: Hare and Lynx Data (yearly data: 1845-1935)

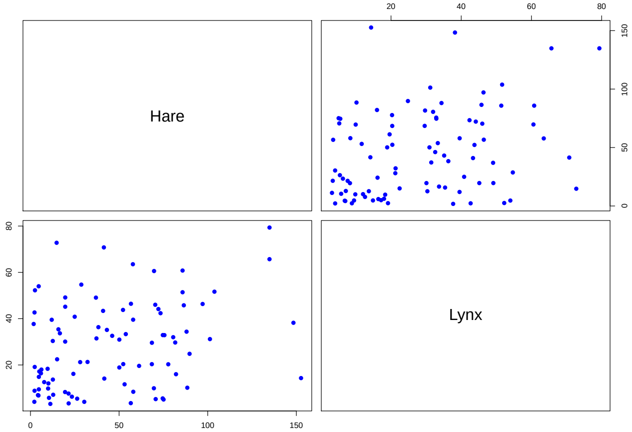


# Example 3: Hare and Lynx Data (cont.)




(1) Increase of pop'n in hare promotes the growth of lynx's pop'n with effects maximized in 1-2 years delay; (2) Increase of pop'n in lynx suppresses the growth of hare's pop'n with effects maximized in 3-4 years delay.

# Example 3: Hare and Lynx Data (cont.)




Relationship between 2 series might not be inspected in a pairwise multivariate setting!

## Dependence Measures in Multiple Time Series

- ACF matrix for 2-dim $\boldsymbol{X}_t$:

$$\underbrace{\Gamma(h)}_{2\times 2} \equiv [\gamma_{ij}(h)] \equiv E[(\boldsymbol{X}_t - \boldsymbol{\mu})(\boldsymbol{X}_{t-h} - \boldsymbol{\mu})'] = \begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{bmatrix} = \Gamma'(-h)$$

- Cross covariance function (CCF):

$$\begin{aligned} \gamma_{12}(h) &\equiv cov(X_{1,t}, X_{2,t-h}), \quad \text{(future } X_1\text{, past } X_2\text{)} \\ \gamma_{21}(h) &= \gamma_{12}(-h) \neq \gamma_{12}(h) = \gamma_{21}(-h). \quad \text{(not symmetric)} \end{aligned}$$

- Cross correlation matrix:

$$\boldsymbol{\rho}(h) \equiv D^{-1/2}\,\Gamma(h)\,D^{-1/2} = \boldsymbol{\rho}'(-h), \text{ where } D = \begin{bmatrix} \gamma_{11}(0) & 0 \\ 0 & \gamma_{22}(0) \end{bmatrix}$$

- Sample CCF for $k$ series:

$$\begin{aligned} \underbrace{\hat{\Gamma}(h)}_{k\times k} &= \frac{1}{T-1}\sum_{t=h+1}^{T} (\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_{t-h} - \bar{\boldsymbol{X}})', \\ \hat{\boldsymbol{\rho}}(h) &= \hat{D}^{-1/2}\,\hat{\Gamma}(h)\,\hat{D}^{-1/2}, \quad \hat{D} = \begin{bmatrix} \hat{\gamma}_{11}(0) & 0 \\ 0 & \hat{\gamma}_{22}(0) \end{bmatrix} \end{aligned}$$

# Test for Serial Dependence

$$H_0 \quad : \quad \boldsymbol{\rho}(1) = \boldsymbol{\rho}(2) = \cdots = \boldsymbol{\rho}(m) = \mathbf{0}_{k\times k}, \quad \text{for some positive integer } m.$$

$$\begin{aligned} Q_k(m) &= T^2 \sum_{h=1}^{m} \frac{1}{T-h} \text{tr}\left[\hat{\boldsymbol{\Gamma}}'(h)\hat{\boldsymbol{\Gamma}}^{-1}(0)\hat{\boldsymbol{\Gamma}}(h)\hat{\boldsymbol{\Gamma}}^{-1}(0)\right] \\ &= T^2 \sum_{h=1}^{m} \frac{1}{T-h} \text{tr}\left[\hat{\boldsymbol{\rho}}'(h)\hat{\boldsymbol{\rho}}^{-1}(0)\hat{\boldsymbol{\rho}}(h)\hat{\boldsymbol{\rho}}^{-1}(0)\right] \\ &= T^2 \sum_{h=1}^{m} \frac{1}{T-h} \left[\text{vec}(\hat{\boldsymbol{\rho}}(h))\right]' \left[\hat{\boldsymbol{\rho}}^{-1}(0) \otimes \hat{\boldsymbol{\rho}}^{-1}(0)\right] \underbrace{\text{vec}(\hat{\boldsymbol{\rho}}(h))}_{k^2\times 1}. \end{aligned}$$

- Under $H_0$ and Gaussian assumption, $Q_k(m) \to \chi^2_{k^2 m}$.
- Special case with $k=1$: $Q_1(m) = T^2 \sum_{h=1}^{m} \hat{\rho}_{11}^2(h)/(T-h) \to \chi^2_m$ reduces to Ljung-Box test statistics
- Note: $\gamma_{ij}(h) = 0 \not\Leftrightarrow \gamma_{ji}(h) = 0$

  (i.e., causality relationship might have a certain direction!)

# $k$-dim Vector ARMA Models

$$(I_k - \Phi_1 B - \cdots - \Phi_p B^p) \underbrace{(\boldsymbol{x}_t - \boldsymbol{\mu})}_{k} = (I_k + \Theta_1 B + \cdots + \Theta_q B^q)\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim WN(\mathbf{0}, \Sigma_\epsilon).$$

- $I_k$, $\Phi_i$'s and $\Theta_i$'s: $k \times k$ matrices
- number of parameters: $\dim(\boldsymbol{\mu}) = k$, $\dim(\Phi_i' s) = k^2 p$, $\dim(\Theta_i' s) = k^2 q$, $\dim(\Sigma_\epsilon) = k(k+1)/2$
- VARMA has a more complex model identifiability problem (details refer to Tsay MTS Book ch3-4)
- LS and ML estimation mathods can both applied (under an identified model specification)
- model (subset) selection becomes crucial since the number of parameters grows very fast as AR and MA orders increase.
- some R functions still work for vector time series, e.g., ar, acf, pacf.
- VAR model can be implemented using R packages: vars, MTS, mAr

## Simplest Case: $k$-dim VAR(1)

$$(I_k - \Phi_1 B)(\boldsymbol{x}_t - \boldsymbol{\mu}) = \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim (\mathbf{0}, \Sigma_\epsilon).$$

- MA representation: $\boldsymbol{x}_t = \boldsymbol{\mu} + (I_k - \Phi_1 B)^{-1}\boldsymbol{\epsilon}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Phi_1^j \boldsymbol{\epsilon}_{t-j}$, provided that the roots of $\det(I_k - \Phi_1 z) = 0$ are outside the unit circle.
- $\Gamma(0) \equiv \mathrm{var}(\boldsymbol{x}_t) = \Phi_1 \Gamma(0) \Phi_1' + \Sigma_\epsilon$, therefore

$$\begin{aligned} & \mathrm{vec}(\Gamma(0)) = \mathrm{vec}(\Phi_1\Gamma(0)\Phi_1') + \mathrm{vec}(\Sigma_\epsilon) = (\Phi_1 \otimes \Phi_1)\mathrm{vec}(\Gamma(0)) + \mathrm{vec}(\Sigma_\epsilon) \\ \Rightarrow \quad & (I_{k^2} - \Phi_1 \otimes \Phi_1)\mathrm{vec}(\Gamma(0)) = \mathrm{vec}(\Sigma_\epsilon) \\ \Rightarrow \quad & \mathrm{vec}(\Gamma(0)) = (I_{k^2} - \Phi_1 \otimes \Phi_1)^{-1}\mathrm{vec}(\Sigma_\epsilon). \\ & \text{special case } k = 1: \quad \gamma(0) = \frac{\sigma_\epsilon^2}{1 - \phi^2}. \end{aligned}$$

- $\Gamma(h) = \mathrm{cov}(\boldsymbol{x}_t, \boldsymbol{x}_{t-h}) = \Phi_1\Gamma(h-1) = \Phi_1^h\,\Gamma(0), \quad$ for $\;h = 1, 2, \ldots.$

## $k$-dim VAR(p)

$$(I - \Phi_1 B - \cdots - \Phi_p B^p)(\boldsymbol{x}_t - \boldsymbol{\mu}) = \boldsymbol{\epsilon}_t$$

is equivalent to

$$\boldsymbol{x}_t = \boldsymbol{v} + \Phi_1 \boldsymbol{x}_{t-1} + \cdots + \Phi_p \boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{v} = \left( I_k - \sum_{j=1}^{p} \Phi_j \right) \boldsymbol{\mu}.$$

- VAR(1) representation: $\boldsymbol{Y}_t = (\mathbf{1}_p \otimes \boldsymbol{v}) + \boldsymbol{\Phi} \boldsymbol{Y}_{t-1} + \boldsymbol{U}_t$, where

$$\boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t-1} \\ \vdots \\ \boldsymbol{x}_{t-p+1} \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \cdots & \Phi_p \\ I_k & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & I_k & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & I_k & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{U}_t = \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$$

- stability condition: all roots of $\det \left( I_k - \sum_{j=1}^{p} \Phi_j z^j \right) = 0$ are outside the unit circle

# Regression Representation for VAR(p)

Let

$$
\begin{aligned}
\boldsymbol{Y} &= (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T), \quad k \times T \\
\boldsymbol{B} &= (\boldsymbol{v}, \Phi_1, \Phi_2, \ldots, \Phi_p), \quad k \times (kp+1) \\
\boldsymbol{z}_t &= (1, \boldsymbol{x}_t', \ldots, \boldsymbol{x}_{t-p+1}')', \quad (kp+1) \times 1 \\
\boldsymbol{Z} &= (\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{T-1}), \quad (kp+1) \times T \\
\boldsymbol{U} &= (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T), \quad k \times T.
\end{aligned}
$$

Then, $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{U}$ and its vector form satisfies the regression:

$$
\begin{aligned}
\boldsymbol{y}^* &\equiv \text{vec}(\boldsymbol{Y}) = \text{vec}(\boldsymbol{B}\boldsymbol{Z}) + \text{vec}(\boldsymbol{U}) \\
&= (\boldsymbol{Z}' \otimes I_k)\, \text{vec}(\boldsymbol{B}) + \text{vec}(\boldsymbol{U}) \equiv \boldsymbol{Z}^* \boldsymbol{\beta}^* + \boldsymbol{u}^*, \\
\text{var}(\boldsymbol{u}^*) &\equiv \Sigma_u = \underbrace{\text{Bdiag}(\Sigma_\epsilon, \ldots, \Sigma_\epsilon)}_{\text{block diagonal matrix}} = I_T \otimes \Sigma_\epsilon.
\end{aligned}
$$

## Least Squares Estimation

$$\boldsymbol{y}^* = (\boldsymbol{Z}' \otimes I_k)\boldsymbol{\beta}^* + \boldsymbol{u}^*, \quad \boldsymbol{u}^* \sim N(\boldsymbol{0}, I_T \otimes \Sigma_\epsilon).$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^*_{GLS} &\equiv \arg\min_\beta \left(\boldsymbol{y}^* - (\boldsymbol{Z}' \otimes I_k)\boldsymbol{\beta}^*\right)'(I_T \otimes \Sigma_\epsilon)^{-1}\left(\boldsymbol{y}^* - (\boldsymbol{Z}' \otimes I_k)\boldsymbol{\beta}^*\right) \\
&= \arg\min_\beta \operatorname{tr}\left[(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{Z})'\Sigma_\epsilon^{-1}(\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{Z})\right] \\
&= \left((\boldsymbol{Z}\boldsymbol{Z}') \otimes \Sigma_\epsilon^{-1}\right)^{-1} (\boldsymbol{Z} \otimes \Sigma_\epsilon^{-1})\boldsymbol{y}^* \\
&= \left((\boldsymbol{Z}\boldsymbol{Z}')^{-1} \otimes \Sigma_\epsilon\right) (\boldsymbol{Z} \otimes \Sigma_\epsilon^{-1})\boldsymbol{y}^* \\
&= \left((\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z} \otimes I_k\right) \boldsymbol{y}^*, \qquad \text{(surprisingly, not dependent on } \Sigma_\epsilon\text{)} \\
\hat{\Sigma}_\epsilon &= \hat{\boldsymbol{U}}\hat{\boldsymbol{U}}'/T = (\boldsymbol{Y} - \hat{\boldsymbol{B}}\boldsymbol{Z})(\boldsymbol{Y} - \hat{\boldsymbol{B}}\boldsymbol{Z})'/T.
\end{aligned}$$

- By asymptotic theory,

$$\begin{aligned}
\sqrt{T}(\hat{\boldsymbol{\beta}}^*_{GLS} - \boldsymbol{\beta}) &\to N(\boldsymbol{0}, \Gamma^{-1} \otimes \Sigma_\epsilon), \\
\hat{\Sigma}_\epsilon &\to \Sigma_\epsilon,
\end{aligned}$$

  where $\Gamma = \lim_{T\to\infty} \frac{1}{T}\boldsymbol{Z}\boldsymbol{Z}'$.

- Again, GLS, OLS and MLE for VAR($p$) are asymptotic equivalent.

# Granger Causality

- Define Notation: PMSE $(z_{t+h}|\mathcal{F}_t)$ denotes the optimal (minimal) MSE for predicting $z_{t+h}$ given information set $\mathcal{F}_t$.
- $\{x_t\}$ is said to cause $\{z_t\}$ in Granger's sense if

  $$\text{PMSE}\left(z_{t+h}|\mathcal{F}_t\right) < \text{PMSE}\left(z_{t+h}|\mathcal{F}_t \setminus \{x_s : s \leq t\}\right), \quad \text{for at least one } h = 1, 2, \ldots.$$

  Also described as $\{x_t\}$ is Granger-causal for $\{z_t\}$. (real effect on forecasting ahead)
- instantaneous causality between $z_t$ and $x_t$ if

  $$\text{PMSE}\left(z_{t+1}|\mathcal{F}_t \cup \{x_{t+1}\}\right) \neq \text{PMSE}\left(z_{t+1}|\mathcal{F}_t\right).$$

  (real effect on nowcasting)

# Granger Causality in VAR Framework

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{v} + \Phi_1 \boldsymbol{x}_{t-1} + \cdots + \Phi_p \boldsymbol{x}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim (\boldsymbol{0}, \Sigma_\epsilon), \\ x_{it} &: \text{$i$-th series in } \boldsymbol{x}_t \\ \Phi_{j,ik} &: \text{the } (i,k) \text{ entry in } \Phi_j \\ \Sigma_{\epsilon,ik} &: \text{the } (i,k) \text{ entry in } \Sigma_\epsilon \end{aligned}$$

- $\{x_{kt}\}$ causes $\{x_{it}\}$ in Granger's sense if

$$\Phi_{j,ik} \neq 0, \quad \text{for some } j = 1, 2, ..., p.$$

- There exists the instantaneous causality between $x_{it}$ and $x_{kt}$ if

$$\Sigma_{\epsilon,ik} \neq 0.$$

- Both parameter statements can be examined via hypotheses testing using likelihood-ratio test, LM or Wald tests.

# Impulse Response Function (IRF)

- Impulse response refers to the reaction of any dynamic system in response to some external change.
- In a VAR model setting, IRF can be characterized via the VMA($\infty$) representation:

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j} = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{U}' \boldsymbol{a}_{t-j},$$

  where var$(\boldsymbol{\epsilon}_t) = \Sigma_\epsilon = \boldsymbol{U}'\boldsymbol{U}$ and $\boldsymbol{a}_t \equiv (\boldsymbol{U}')^{-1}\boldsymbol{\epsilon}_t \sim (\boldsymbol{0}, \boldsymbol{I})$.
- $\text{IRF}_j = \frac{\partial \boldsymbol{x}_{t+j}}{\partial \boldsymbol{\epsilon}_t} = \boldsymbol{\Psi}_j$
  - $\text{IRF}_j(i, \ell)$ indicates the impact on $i$-th variable due to the external change of $\ell$-th variable with time lag $j$.
  - external change in $\boldsymbol{\epsilon}_t$ is cross-correlated and therefore cross reactions may occur
- Orthogonal $\text{IRF}_j = \frac{\partial \boldsymbol{x}_{t+j}}{\partial \boldsymbol{a}_t} = \boldsymbol{\Psi}_j \boldsymbol{U}'$
  - external change in $\boldsymbol{a}_t$ is orthogonal; cross reactions are eliminated

# R Demo: cmort effected by PM

External shocks of PM reach the maximum effect on **cmort** after 8-10 weeks on time.

# Transfer Function Model (Lag Regression)

- Notation: $\{x_t\}$ is the input series, $\{y_t\}$ is the output series.
- Aim: modeling the relationship only in one-direction between two series (mainly interested in $y_t$).
- Model specification:

$$\begin{aligned} y_t &= \alpha(B)x_t + \eta_t, \quad \alpha(B) = \sum_{j=0}^{\infty} \alpha_j B^j, \\ \phi(B)x_t &= \theta(B)w_t, \quad w_t \sim WN(0, \sigma_w^2), \end{aligned}$$

  where $\{\eta_t\}$ is a stationary process and uncorrelated with $\{x_t\}$ (which characterizes the dynamics of $y_t$ leftover after explained by $x_t$).
- Fitting procedures:
  1. $\gamma_{yx}(k) = \sum_{j=0}^{\infty} \alpha_j \gamma_{xx}(k-j)$ satisfying $\alpha_j = \gamma_{yx}(j)/\gamma_{xx}(0)$ if $\{x_t\} \sim WN$.
  2. pre-whitening $\{x_t\}$ to get $\{w_t\}$: i.e., estimate $\phi(B)$ and $\theta(B)$ and define $\pi(B) = \theta^{-1}(B)\phi(B)$. Let

$$y_t^* \equiv \pi(B)y_t = \alpha(B)[\pi(B)x_t] + \underline{\pi(B)\eta_t} = \alpha(B)w_t + \underline{\eta_t^*},$$

     which forms a new transfer function model with same $\alpha(B)$ but the input series $\{w_t\}$ becomes WN.

## Transfer Function Model: Fitting Procedures (cont.)

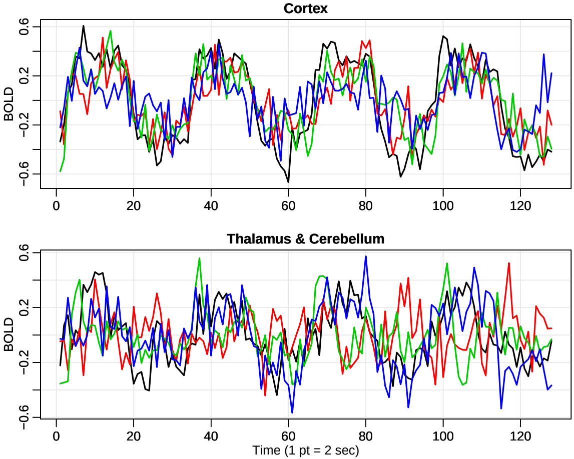
3. estimate $\alpha(B)$ initially by $\hat{\alpha}_j = \hat{\gamma}_{y^*w}(j)/\hat{\gamma}_{ww}(0)$. Try to approximate $\hat{\alpha}(B)$ by a simpler structure such as $B^d\delta_s(B)/\beta_r(B)$ where $\delta_s(B) = \sum_{j=0}^{s}\delta_j B^j$ ($\delta_0$ may not be 1) and $\beta_r(B) = 1 - \sum_{j=1}^{r}\beta_j B^j$. Determine the nonnegative integer $d$ (leading lag).
4. back to the original model:

$$
\begin{aligned}
y_t &= \alpha(B)x_t + \eta_t = B^d\delta_s(B)/\beta_r(B)x_t + \eta_t, \\
\beta_r(B)y_t &= B^d\delta_s(B)x_t + \beta_r(B)\eta_t, \\
y_t &= \beta_1 y_{t-1} + \cdots + \beta_r y_{t-r} + \delta_0 x_{t-d} + \cdots + \delta_s x_{t-d-s} + u_t,
\end{aligned}
$$

   where $u_t \equiv \beta_r(B)\eta_t$. Estimate $(\beta_1, \ldots, \beta_r)'$ and $(\delta_0, \ldots, \delta_s)'$ by regression.
5. obtain the regression residuals $\{\hat{u}_t\}$ and solve $\hat{\eta}_t = \beta_r^{-1}(B)\hat{u}_t$. Fit an ARMA to $\hat{\eta}_t$. Or, fit $\{u_t\}$ directly.

## Common Patterns in High-dim Time Series

fMRI data (blood oxygenation-level dependent signal intensity) from various locations in the cortex, thalamus and cerebellum



# (Dynamic) Factor Model for High-dim Time Series

$$\boldsymbol{Y}_t \quad = \quad \underbrace{\boldsymbol{X}_t\boldsymbol{\beta}}_{\text{determinic term}} \quad + \quad \underbrace{A\boldsymbol{f}_t}_{\text{dynamic term}} \quad + \quad \underbrace{\boldsymbol{\epsilon}_t}_{\text{error term}} \qquad (*)$$

- $\{\boldsymbol{Y}_t\}$ are observable multivariate data among which the dependence is driven by lower-dimensional (common) dynamic terms in $\{\boldsymbol{f}_t\}$, i.e., $\dim(\boldsymbol{f}_t) << \dim(\boldsymbol{Y}_t)$ (dimension reduction)
- $\{\boldsymbol{f}_t\}$ are unobservable and typically assumed following a VAR model, e.g.,

$$\boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \quad (**)$$

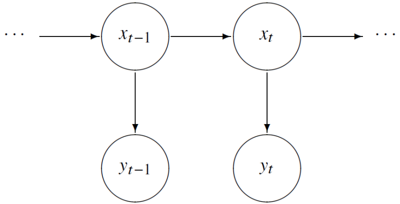
- $(*) - (**)$ called a linear state-space model
- Inference Goal: infer the dynamics of common factors $\boldsymbol{f}_t$ and predict both $\boldsymbol{f}_t$ and $\boldsymbol{Y}_t$ in advance.

# State-Space Models (SSM)

A simple example:

$$y_t = x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$
$$x_t = \phi x_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

where $\eta_t$ is independent of $\{\epsilon_t\}$.




- special case with $\phi = 1 \;\Leftrightarrow\;$ random level model (random walk + WN)
- observe $\{y_t\}$; but interested in the unobserved signal $\{x_t\}$
- Goal: predict $\{x_t\}$ given $\{y_t\}$
- How to make inference?

  Kalman filter provides on-line forecasting

# Linear State-Space Models

General form:

$$\text{observation equation:} \qquad \boldsymbol{y}_t = Z_t\boldsymbol{\alpha}_t + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, H_t),$$

$$\text{state equation:} \qquad \boldsymbol{\alpha}_t = T_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + R_t\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, Q_t).$$

where $\{\boldsymbol{\epsilon}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are serially uncorrected and mutually independent.

- system matrices $Z_t, \boldsymbol{d}_t, H_t, T_t, \boldsymbol{c}_t, R_t, Q_t$ are assumed to be non-stochastic for simplicity (could be stochastic in general)
- the system is time-invariant (time-homogeneous) if the system matrices do not depend on time
- initial state $\boldsymbol{\alpha}_0$ could be a fixed point, e.g., the marginal mean; or assumed as $\boldsymbol{\alpha}_0 \sim N(\boldsymbol{a}_0, P_0)$, e.g., the marginal distribution of $\boldsymbol{a}_t$

# More Examples of SSM

- transformed stochastic volatility (SV) model:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t, \\ \underbrace{\log r_t^2}_{\text{obs'n}} &= \log \sigma_t^2 + \log \epsilon_t^2 = \mu + \underbrace{v_t}_{\text{state}} + \epsilon_t^* \end{aligned}$$

- local linear trend with random slope (Harvey, 1993):

$$y_t = \mu_t + \epsilon_t, \quad \mu_t = \mu_{t-1} + \beta_{t-1} + \xi_t, \quad \beta_t = \beta_{t-1} + \eta_t,$$

can be written as

$$\begin{aligned} y_t &= (1,0) \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \epsilon_t, \\ \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix}. \end{aligned}$$

- AR($p$), MA($q$) and ARMA($p$,$q$) all have SSM representations

# Inference for SSM

Define

Info set: $Y_{t-1} = (y_1, y_2, \ldots, y_{t-1})'$.

- prediction problem: predict $\boldsymbol{\alpha}_t$ given $Y_{t-1}$

  compute $\boldsymbol{a}_{t|t-1} = E[\boldsymbol{\alpha}_t | Y_{t-1}]$; $P_{t|t-1} \equiv E(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})'$

- filtering problem: predict $\boldsymbol{\alpha}_t$ given $Y_t$

  compute $\boldsymbol{a}_t \equiv \boldsymbol{a}_{t|t} = E[\boldsymbol{\alpha}_t | Y_t]$; $P_t \equiv P_{t|t} = E(\boldsymbol{\alpha}_t - \boldsymbol{a}_t)(\boldsymbol{\alpha}_t - \boldsymbol{a}_t)'$

- smoothing problem: predict $\boldsymbol{\alpha}_t$ given $Y_n$

  compute $\boldsymbol{a}_{t|n} = E[\boldsymbol{\alpha}_t | Y_n]$; $P_{t|n} \equiv E(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|n})(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|n})'$

# Kalman Filter (KF)

KF is an algorithm to compute $\boldsymbol{a}_{t|t-1}, \boldsymbol{a}_t, P_{t|t-1}, P_t,$ recursively for linear Gaussian SSM, including two sets of equations:

- prediction equations:

$$
\begin{aligned}
\boldsymbol{a}_{t|t-1} &= T_t \boldsymbol{a}_{t-1} + \boldsymbol{c}_t, \\
P_{t|t-1} &= T_t P_{t-1} T_t' + R_t Q_t R_t', \\
\hat{\boldsymbol{y}}_{t|t-1} &= Z_t \boldsymbol{a}_{t|t-1} + \boldsymbol{d}_t, \\
\boldsymbol{v}_t &= \boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}, \\
F_t &= var(\boldsymbol{v}_t) = Z_t P_{t|t-1} Z_t' + H_t.
\end{aligned}
$$

- updating equations:

$$
\begin{aligned}
\boldsymbol{a}_t &= \boldsymbol{a}_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} \boldsymbol{v}_t, \\
P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}.
\end{aligned}
$$

- initial conditions: $\boldsymbol{a}_0 = E\boldsymbol{\alpha}_0$, $P_0 = P_{0|0} = E\left[(\boldsymbol{\alpha_0} - \boldsymbol{a}_0)'(\boldsymbol{\alpha_0} - \boldsymbol{a}_0)\right] = var(\boldsymbol{\alpha}_0)$

# Derivation for Kalman Filter (1)

$$\text{observation equation:} \quad \boldsymbol{y}_t = Z_t\boldsymbol{\alpha}_t + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, H_t),$$

$$\text{state equation:} \quad \boldsymbol{\alpha}_t = T_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + R_t\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, Q_t).$$

- Given $\boldsymbol{a}_{t-1} = E[\boldsymbol{\alpha}_{t-1}|Y_{t-1}]$ and $P_{t-1} = E(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})'$, taking expectation on state eq:

$$\begin{aligned}
\boldsymbol{a}_{t|t-1} &\equiv E[\boldsymbol{\alpha}_t|Y_{t-1}] = T_t \underbrace{E[\boldsymbol{\alpha}_{t-1}|Y_{t-1}]}_{\boldsymbol{a}_{t-1}} + \boldsymbol{c}_t + R_t \underbrace{E[\boldsymbol{\eta}_t|Y_{t-1}]}_{0}, \\
P_{t|t-1} &\equiv E[(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})'|\mathbf{Y}_{t-1}] \\
&= E\left[(T_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + R_t\boldsymbol{\eta}_t) - (T_t\boldsymbol{a}_{t-1} + \boldsymbol{c}_t)\right]\left[\text{same as the first ()}\right]' \\
&= E[T_t\underline{(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})} + R_t\boldsymbol{\eta}_t][T_t\underline{(\boldsymbol{\alpha}_{t-1} - \boldsymbol{a}_{t-1})} + R_t\boldsymbol{\eta}_t]' \\
&= T_tP_tT_t' + R_tQ_tR_t'.
\end{aligned}$$

# Derivation for Kalman Filter (2)

$$\text{observation equation:} \quad \boldsymbol{y}_t = Z_t\boldsymbol{\alpha}_t + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, H_t),$$

$$\text{state equation:} \quad \boldsymbol{\alpha}_t = T_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + R_t\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\boldsymbol{0}, Q_t).$$

- Given $\boldsymbol{a}_{t|t-1}$ and $P_{t|t-1}$, taking expectation on measurement eq:

$$\begin{aligned}
\hat{\boldsymbol{y}}_{t|t-1} &\equiv E[\boldsymbol{y}_t|\boldsymbol{Y}_{t-1}] = Z_t \underbrace{E[\boldsymbol{\alpha}_t|\boldsymbol{Y}_{t-1}]}_{\boldsymbol{a}_{t|t-1}} + \boldsymbol{d}_t + \underbrace{E[\boldsymbol{\epsilon}_t|\boldsymbol{Y}_{t-1}]}_{0}, \\
F_t &\equiv E\Big(\underbrace{\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}}_{\boldsymbol{v}_t}\Big)\Big(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1}\Big)' = var(\boldsymbol{v}_t) \\
&= E\Big((Z_t\boldsymbol{\alpha}_t + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t) - (Z_t\boldsymbol{a}_{t|t-1} + \boldsymbol{d}_t)\Big)\Big(\text{same as the first ()}\Big)' \\
&= E\Big(Z_t\underline{(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})} + \boldsymbol{\epsilon}_t\Big)\Big(\text{same as the first ()}\Big)' \\
&= Z_t\underline{P_{t|t-1}}Z_t' + H_t.
\end{aligned}$$

# Derivation for Kalman Filter (3)

Given $\boldsymbol{Y}_{t-1}$,

$$
\begin{aligned}
\left( \begin{array}{c} \boldsymbol{y}_t \\ \boldsymbol{\alpha}_t \end{array} \right) \Bigg|_{\boldsymbol{Y}_{t-1}} \quad &\sim \quad N\left( \left( \begin{array}{c} \hat{\boldsymbol{y}}_{t|t-1} \\ \boldsymbol{a}_{t|t-1} \end{array} \right), \left( \begin{array}{cc} F_t & Z_t P_{t|t-1} \\ P_{t|t-1} Z_t' & P_{t|t-1} \end{array} \right) \right), \qquad (*) \\
\text{cov}\left(\boldsymbol{y}_t, \boldsymbol{\alpha}_t \middle| \boldsymbol{Y}_{t-1}\right) \quad &= \quad E\left[ (\boldsymbol{y}_t - \hat{y}_{t|t-1})(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1}) \middle| \boldsymbol{Y}_{t-1} \right] \\
&= \quad E\left[ (Z_t \underline{(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})} + \boldsymbol{\epsilon}_t) \underline{(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1})'} \middle| \boldsymbol{Y}_{t-1} \right] = Z_t \underline{P_{t|t-1}},
\end{aligned}
$$

By the results of conditional normal distribution, $(*)$ implies

$$
\begin{aligned}
\boldsymbol{a}_t \quad &\equiv \quad E[\boldsymbol{\alpha}_t | \boldsymbol{Y}_{t-1}, \boldsymbol{y}_t] = \boldsymbol{a}_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} \underbrace{(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1})}_{\boldsymbol{v}_t}, \\
P_t \quad &= \quad \text{var}(\boldsymbol{\alpha}_t | \boldsymbol{Y}_{t-1}, \boldsymbol{y}_t) = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}.
\end{aligned}
$$

# Likelihood Function for SSM

KF produces the innovations $\boldsymbol{v}_t's$ (one-step prediction error) and their variances $F_t's$ (one-step PMSE), so that the log-likelihood (upto a constant) can be represented as

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \underbrace{f(\boldsymbol{y}_t|\mathbf{Y}_{t-1})}_{\sim N\left(\hat{\boldsymbol{y}}_{t|t-1}, F_t\right)} = -\frac{1}{2}\sum_{t=1}^{n}\log|F_t| - \frac{1}{2}\sum_{t=1}^{n}\underbrace{(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1})'}_{\boldsymbol{v}_t'} F_t^{-1} \underbrace{(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_{t|t-1})}_{\boldsymbol{v}_t}.$$

- $\boldsymbol{v}_t$ and $F_t$ are implicit functions of $\boldsymbol{\theta}$ (middle products of KF)
- $\ell(\boldsymbol{\theta})$ can be computed by running KF given parameter value $\boldsymbol{\theta}$
- MLE of $\boldsymbol{\theta}$ can be numerically solved by minimizing $-\log \ell(\boldsymbol{\theta})$
  (write an objective function, to be minimized, with input $\boldsymbol{\theta}$ and output $-\log \ell(\boldsymbol{\theta})$)
- solve MLE of $\boldsymbol{\theta}$ via KF is equivalent to EM (E-step: impute missing states $\boldsymbol{\alpha}_t$'s)
- related functions in R: Kfilter(), Ksmooth()
- related R packages:
  dlm (dynamic linear model); MARSS (multivariate auto-regressive(1) state-space)

# Smoothing Algorithm

- on-line:
  - fixed point smoothing: find $\boldsymbol{a}_{t|n}$ for fixed $t$ as $n$ increases
  - fixed lag smoothing: find $\boldsymbol{a}_{n-\tau|n}$ for fixed $\tau$ as $n$ increases
- off-line:
  - fixed interval smoothing: find $\boldsymbol{a}_{t|n}$ for all $t \leq n$ and fixed $n$ (we focus on this)

# Fixed Interval Smoothing

- Note that the filtered value will be equal to the smoother at $t = n$, i.e.,

$$\boldsymbol{a}_{n|n} = E[\boldsymbol{\alpha}_n | \boldsymbol{Y}_n] = \boldsymbol{a}_n, \quad P_{n|n} = P_n. \quad \text{(initial values)}$$

- Idea: run KF forward to time $n$, then work backward to get smoother.
- smoothing algorithm:

$$\begin{aligned}
P_t^* &= P_t T_{t+1}' P_{t+1|t}^{-1}, \\
\boldsymbol{a}_{t|n} &= \boldsymbol{a}_t + P_t^* (\boldsymbol{a}_{t+1|n} - \boldsymbol{a}_{t+1|t}), \\
P_{t|n} &= P_t - P_t^* \underbrace{(P_{t+1|t} - P_{t+1|n})}_{\text{positive definite}} (P_t^*),'
\end{aligned}$$

for $t = n-1, n-2, \ldots, 1$.

# R Demo: SSM for Temperature Data

Extract a common signal from multiple data source:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} x_t + \underbrace{\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\epsilon}_t}, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}),$$

$$x_t = \underbrace{\delta}_{\text{long-term drift}} + x_{t-1} + \eta_t, \quad w_t \sim N(0, \sigma_w^2).$$

- Analysis detailed in gtemp.html
- Similar problems:
  - Integrating $PM_{2.5}$ data collected from meteorological station (more accurate and sparse in space) and from Airbox (less accurate but very dense in space) to enhance the spatial prediction locally .
  - hidden markov model: hidden inputs multiple outputs problem
    summarizing a hidden factor (common factor) which drives the dynamics of multiple return series.

# R Demo: SSM for Target tracking

Target tracking task for the true location of object, taking into account measurement error ($\boldsymbol{\epsilon}_t$) in observed location and time-varying speed ($\boldsymbol{\beta}_t$):

$$\text{true location:} \quad \boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_\eta),$$

$$\text{speed:} \quad \boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, \;\; \boldsymbol{w}_t \sim N(0, \Sigma_w).$$

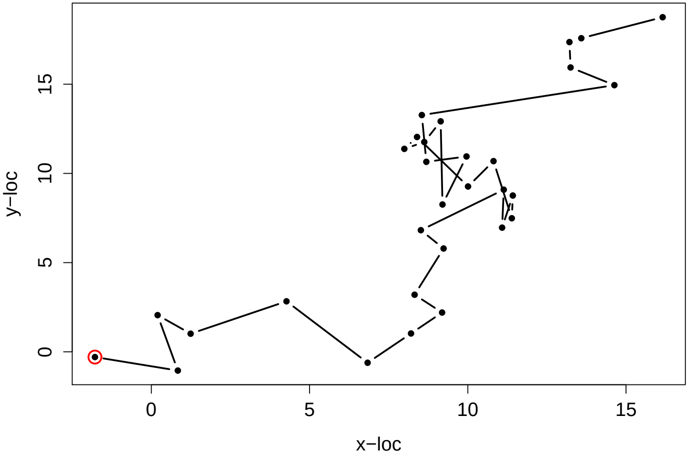
SSM:

$$\text{measurement eq:} \quad \boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t = (I, \mathbf{0})\boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \;\; \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}),$$

$$\text{state eq:} \quad \boldsymbol{x}_t \equiv \begin{pmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\beta}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_{t-1} \\ \boldsymbol{\beta}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{w}_t \end{pmatrix}.$$

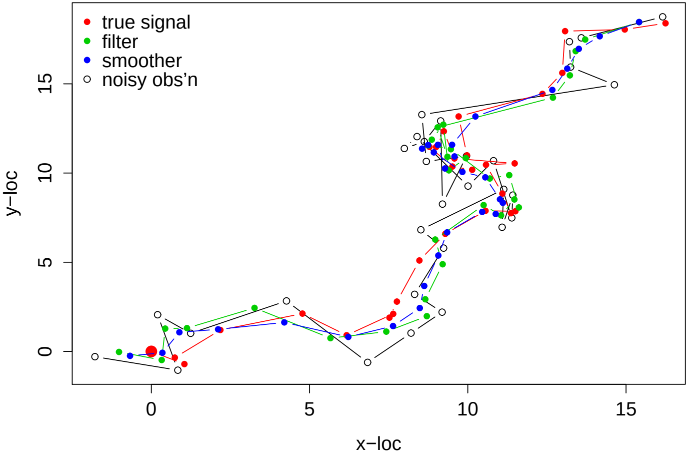
- 2-dim version of the local linear trend model with random slope
- Analysis detailed in tracking.html

# Demo: Turtle's Migration Path (2-dim RW+WN)




Spatial locations along with the path are recorded over time

# Target Tracking: Filtering and Smoothing (2-dim RW+WN)



# Other Topics in Multiple Time Series

- multivariate volatility model
- co-integration:

  looking for predictable combinations of series

  long-term balance relationship among nonstationary multiple series
- nonlinear modeling: toward to local stationary models and network models

## Nonlinear Processes (extensions from AR)

- Regime switching model:

$$
\begin{aligned}
y_t &= \phi_1^{(I_t)} y_{t-1} + \sigma^{(I_t)} \epsilon_t, \quad \epsilon_t \sim N(0,1), \\
\{I_t\} &: \quad \text{discrete-time Markov chain with finite states.}
\end{aligned}
$$

- Threshold AR model:

$$
y_t = \begin{cases} \phi_1 y_{t-1} + \sigma_1 \epsilon_t; & y_{t-1} < 0, \\ \phi_2 y_{t-1} + \phi_3 y_{t-2} + \sigma_2 \epsilon_t; & y_{t-1} \geq 0. \end{cases}
$$

- Random coefficient model:

$$
y_t = \phi_t y_{t-1} + \epsilon_t, \quad \{\phi_t\} \text{ follows some stochastic process,}
$$

  which can be seen as a state-space model.

- Functional coefficient AR model:

$$
\begin{aligned}
y_t &= \phi(\boldsymbol{x}_t) y_{t-1} + \epsilon_t, \quad \phi(\cdot) \text{ is a determinstic function,} \\
\boldsymbol{x}_t &: \quad \text{covariates known at time t, e.g., lag variable } y_{t-j}
\end{aligned}
$$

## Discrete-valued Time Series (via generalized linear model setting)

- Discrete data: e.g.,

$$\begin{aligned} y_t &\sim Ber(p_t); \quad x_t = \log[p_t/(1-p_t)], \\ y_t &\sim Poi(\mu_t); \quad x_t = \log \mu_t, \end{aligned}$$

  - Dynamics assumed on latent process: $\{x_t\} \sim AR$
  - $\{y_t\}$ are commonly assumed to be conditional independent given $\{x_t\}$
  - Such model has a nonlinear and non-Gaussian state space form.
  - Bayesian inference is often used for these types of models.

- Categorical data:

  modeled by a discrete-state Markov chain: specify the transition matrix on finite states, e.g., 3 states of interest in epidemiology:

  $(S_t, I_t, R_t)$: (Susceptible, Infectious, or Recovered).

# Long Memory v.s. Short Memory Processes

- Long-memory (LM) phenomena are commonly observed in finance, economics, geophysics, climate among other fields.
- Definition:
    - ACF decays too slow such that $\sum_h |\gamma(h)| = \infty$
    - spectral density has a pole at frequency zero, i.e., $f(\omega) \sim \omega^{-2d}$
- In contrast, ARMA$(p, q)$ is short memory with exponentially decaying ACF and their sum satisfying $\sum_h |\gamma(h)| < \infty$.
- Most popular LM model:
    - fractionally integrated (FI) process:

$$(1 - B)^d X_t = Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2), \quad 0 < d < 0.5.$$

    - ARFIMA model: $\phi(B)(1 - B)^d X_t = \theta(B) Z_t$
- empirical data show LM phenomenon in volatility structure, related models include FIGARCH...
- ARFIMA can be implemented using R package fracdiff