Assessing nonlinear Granger causality from multivariate time series

Xiaohai Sun

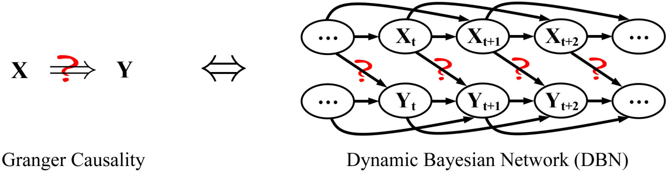
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

September 17, 2008

Learning of causality in time series

- $\mathrm{X} := (\ldots, x_{t-1}, x_t, x_{t+1}, \ldots)$: discrete time, continuous state process
  $t \in \mathbb{Z}$: discrete time point, usually taken at equally spaced intervals
- Time-delayed vector: $X_t := (x_{t-n+1}, \ldots, x_{t-1}, x_t)^\mathsf{T}$
- Task: bivariate time series (X, Y) observed, estimate whether the underlying process of X is causal to the underlying process of Y or/and the other way around.
- Notation: "$\mathrm{X} \quad \mathrm{Y}$", "$\mathrm{X} \Rightarrow \mathrm{Y}$", "$\mathrm{X} \Leftarrow \mathrm{Y}$", or "$\mathrm{X} \Leftrightarrow \mathrm{Y}$"
- Granger's concept of causality (1969): $\mathrm{X} \Rightarrow \mathrm{Y}$, if the future values of Y can be better predicted using the past values of X and Y compared to using the past values of Y alone.

Linear Granger causality



- Standard test: linear autoregression models

$$Y_{t+1} = a^{\mathsf{T}} \cdot Y_t + \epsilon^{(\mathrm{Y})} \quad \text{and} \quad Y_{t+1} = b_1^{\mathsf{T}} \cdot Y_t + b_2^{\mathsf{T}} \cdot X_t + \epsilon^{(\mathrm{Y}|\mathrm{X})},$$

  a, b_1, b_2: regression coefficient vectors, determined so that prediction errors $\mathrm{Var}[\epsilon^{(\mathrm{Y})}]$ and $\mathrm{Var}[\epsilon^{(\mathrm{Y}|\mathrm{X})}]$ minimize.

- If $\mathrm{Var}[\epsilon^{(\mathrm{Y}|\mathrm{X})}] \ll \mathrm{Var}[\epsilon^{(\mathrm{Y})}]$, then $\mathrm{X} \Rightarrow \mathrm{Y}$.

Nonlinear Granger causality

- Our proposal: nonlinear autoregression models

$$\begin{aligned} a^\top \cdot \Phi(Y_{t+1}) &= b_1^\top \cdot \Psi(Y_t) + \epsilon^{(\mathrm{Y})} \\ a^\top \cdot \Phi(Y_{t+1}) &= b_2^\top \cdot \Psi(Y_t, X_t) + \epsilon^{(\mathrm{Y}|\mathrm{X})} \end{aligned}$$

Φ, Ψ: nonlinear maps into some feature spaces.

- If $\mathrm{Var}[\epsilon^{(\mathrm{Y}|\mathrm{X})}] \ll \mathrm{Var}[\epsilon^{(\mathrm{Y})}]$, then $\mathrm{X} \Rightarrow \mathrm{Y}$.
- Extension to conditional cases: $\mathrm{X} \Rightarrow \mathrm{Y} \mid Z$

$$\begin{aligned} a^\top \cdot \Phi(Y_{t+1}) &= b_1^\top \cdot \Psi(Y_t, Z_t) + \epsilon^{(\mathrm{Y})} \\ a^\top \cdot \Phi(Y_{t+1}) &= b_2^\top \cdot \Psi(Y_t, Z_t, X_t) + \epsilon^{(\mathrm{Y}|\mathrm{X})} \end{aligned}$$

Embedding of distributions in RKHS

- $\mathcal{H}_\mathcal{Y}$: Hilbert space on measurable space $\mathcal{Y}$, spanned by functions $k_\mathcal{Y}(y,\cdot)$ $(y \in \mathcal{Y})$ with $\langle k_\mathcal{Y}(y,\cdot), k_\mathcal{Y}(y',\cdot)\rangle = k_\mathcal{Y}(y,y')\ \forall y, y' \in \mathcal{Y}$.
  Y: random variable on $\mathcal{Y}$.

- Mean element in RKHS:
  $\mathfrak{M}_Y = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)]$ and $\mathfrak{M}_{YY} = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)k_\mathcal{Y}(Y,\cdot)]$
- Conditional mean element in RKHS:
  $\mathfrak{M}_{Y|X} = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)|X]$ and $\mathfrak{M}_{YY|X} = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)k_\mathcal{Y}(Y,\cdot)|X]$
- Product of mean elements in RKHS:
  $\mathfrak{M}_Y\mathfrak{M}_Y = \mathfrak{M}_Y \otimes \mathfrak{M}_Y = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)]\mathrm{E}[k_\mathcal{Y}(Y,\cdot)]$
- Product of conditional mean elements in RKHS:
  $\mathfrak{M}_{Y|X}\mathfrak{M}_{Y|X} = \mathfrak{M}_{Y|X} \otimes \mathfrak{M}_{Y|X} = \mathrm{E}[k_\mathcal{Y}(Y,\cdot)|X]\mathrm{E}[k_\mathcal{Y}(Y,\cdot)|X]$

Covariance operator

- Covariance operator in RKHS:

$$\begin{aligned}\langle g, \Sigma_{YY} g \rangle_{\mathcal{H}_\mathcal{Y}} &:= \langle \mathfrak{M}_{YY} - \mathfrak{M}_Y \mathfrak{M}_Y, g \otimes g \rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \\ &= \mathrm{E}[g(Y)g(Y)] - \mathrm{E}[g(Y)]\mathrm{E}[g(Y)] \\ &= \mathrm{Var}[g(Y)] \qquad \forall g \in \mathcal{H}_\mathcal{Y}\end{aligned}$$

- Conditional covariance operator in RKHS:

$$\begin{aligned}\left\langle g, \Sigma_{YY|X} g \right\rangle_{\mathcal{H}_\mathcal{Y}} &:= \left\langle \mathfrak{M}_{YY} - \mathrm{E}_X[\mathfrak{M}_{Y|X} \mathfrak{M}_{Y|X}], g \otimes g \right\rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \\ &= \mathrm{E}[g(Y)g(Y)] - \mathrm{E}_X[\mathrm{E}[g(Y)|X]\mathrm{E}[g(Y)|X]] \\ &= \mathrm{E}_X[\mathrm{Var}[g(Y) \,|\, X]] \qquad \forall g \in \mathcal{H}_\mathcal{Y}\end{aligned}$$

Difference of covariance operator and mean elements

$$
\begin{aligned}
& \langle g, \Sigma_{YY} g \rangle_{\mathcal{H}_\mathcal{X}} - \langle g, \Sigma_{YY|X} g \rangle_{\mathcal{H}_\mathcal{Y}} \\
= & \langle \mathrm{E}_X[\mathfrak{M}_{Y|X} \mathfrak{M}_{Y|X}] - \mathfrak{M}_Y \mathfrak{M}_Y, g \otimes g \rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \\
= & \mathrm{Var}_X[\mathrm{E}_Y[g(Y)|X]] \quad \geq 0 \qquad \forall g \in \mathcal{H}_\mathcal{Y}
\end{aligned}
$$

$$
\begin{aligned}
& \langle g, \Sigma_{YY|Z} g \rangle_{\mathcal{H}_\mathcal{X}} - \langle g, \Sigma_{YY|XZ} g \rangle_{\mathcal{H}_\mathcal{Y}} \\
= & \langle \mathrm{E}_{XZ}[\mathfrak{M}_{Y|XZ} \mathfrak{M}_{Y|XZ}] - \mathrm{E}_Z[\mathfrak{M}_{Y|Z} \mathfrak{M}_{Y|Z}], g \otimes g \rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \\
= & \mathrm{E}_Z[\mathrm{Var}_X[\mathrm{E}_Y[g(Y)|X, Z]]] \quad \geq 0 \qquad \forall g \in \mathcal{H}_\mathcal{Y}
\end{aligned}
$$

Order of mean elements and covariance operators

- Order of mean elements

$$\mathfrak{M}_Y \mathfrak{M}_Y \;\leq\; \mathrm{E}_Z[\mathfrak{M}_{Y|Z} \mathfrak{M}_{Y|Z}] \;\leq\; \mathrm{E}_{XZ}[\mathfrak{M}_{Y|XZ} \mathfrak{M}_{Y|XZ}] \;\leq\; \cdots$$

in the sense, for all $g \otimes g \in \mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}$

$$\langle \mathfrak{M}_Y \mathfrak{M}_Y, g \otimes g \rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \;\leq\; \left\langle \mathrm{E}_Z[\mathfrak{M}_{Y|Z} \mathfrak{M}_{Y|Z}], g \otimes g \right\rangle_{\mathcal{H}_\mathcal{Y} \otimes \mathcal{H}_\mathcal{Y}} \;\leq\; \cdots$$

- Order of covariance operators

$$\cdots \;\leq\; \Sigma_{YY|XZ} \;\leq\; \Sigma_{YY|Z} \;\leq\; \Sigma_{YY} \,,$$

in the sense, for all $g \in \mathcal{H}_\mathcal{Y}$

$$0 \;\leq\; \cdots \;\leq\; \left\langle g, \Sigma_{YY|XZ} g \right\rangle_{\mathcal{H}_\mathcal{Y}} \;\leq\; \left\langle g, \Sigma_{YY|Z} g \right\rangle_{\mathcal{H}_\mathcal{Y}} \;\leq\; \langle g, \Sigma_{YY} g \rangle_{\mathcal{H}_\mathcal{Y}}$$

Significance test of predictability

- Hilbert-Schmidt (HS) norm of operator Σ:

$$\|\Sigma\|_{\mathrm{HS}}^2 = \mathrm{Tr}(\Sigma^\mathsf{T}\Sigma)$$

- Unpredictability by HS norm

$$\begin{aligned} \|\Sigma_{YY|X}\|_{\mathrm{HS}}^2 &= \|\Sigma_{YY}\|_{\mathrm{HS}}^2 &\iff& \quad X \not\Rightarrow Y \\ \|\Sigma_{YY|XZ}\|_{\mathrm{HS}}^2 &= \|\Sigma_{YY|Z}\|_{\mathrm{HS}}^2 &\iff& \quad X \not\Rightarrow Y \mid Z \end{aligned}$$

- Significance test via random permutation π_j:

$$\begin{aligned} \|\Sigma_{YY|X}\|_{\mathrm{HS}}^2 &\ll \|\Sigma_{YY|X^{\pi_j}}\|_{\mathrm{HS}}^2 \approx \|\Sigma_{YY}\|_{\mathrm{HS}}^2 \\ \|\Sigma_{YY|XZ}\|_{\mathrm{HS}}^2 &\ll \|\Sigma_{YY|X^{\pi_j}Z}\|_{\mathrm{HS}}^2 \approx \|\Sigma_{YY|Z}\|_{\mathrm{HS}}^2 \end{aligned}$$

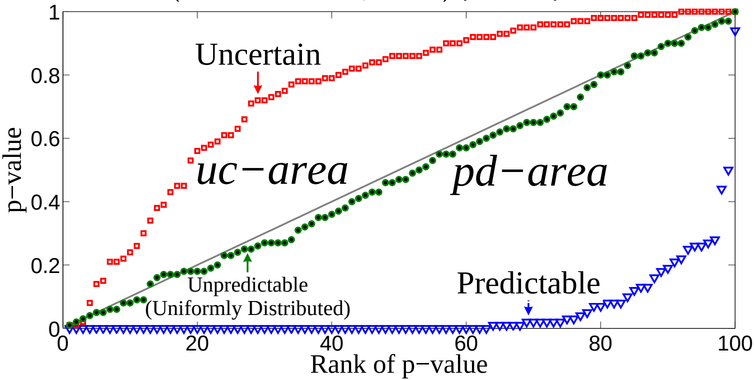
Note: No need to partition conditioning variable Z.

Subsampling-based multiple testing

m sub-time-series with pre-specified size $n_0 \ll n$ $(n \approx m \cdot n_0)$

$$\|\widehat{\Sigma}^{(n_0)}_{YY|X}\|^2_{\mathrm{HS}} \ll \|\widehat{\Sigma}^{(n_0)}_{YY|X^{\pi_j}}\|^2_{\mathrm{HS}} \quad \text{or} \quad \|\widehat{\Sigma}^{(n_0)}_{YY|XZ}\|^2_{\mathrm{HS}} \ll \|\widehat{\Sigma}^{(n_0)}_{YY|X^{\pi_j}Z}\|^2_{\mathrm{HS}}$$

Q-Q ("Q" stands for quantile) plot of p-values

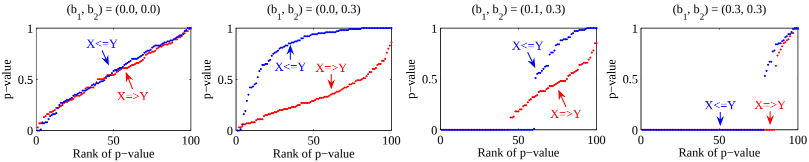


Simulated data: chaotic maps

Noisy logistic maps:
b_1: coupling strength of $\mathrm{X} \Leftarrow \mathrm{Y}$; b_2: coupling strength of $\mathrm{X} \Rightarrow \mathrm{Y}$

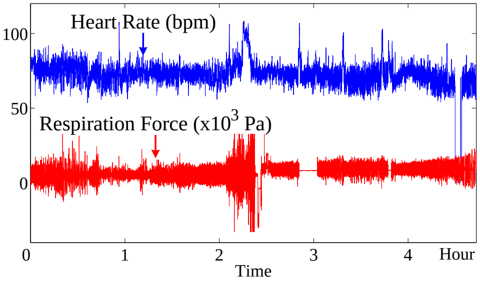
$$
\begin{aligned}
x_{t+1} &= (1-b_1)\, a\, x_t(1-x_t) + b_1\, a\, y_t(1-y_t) + \mu\, \xi_1 \\
y_{t+1} &= (1-b_2)\, a\, y_t(1-y_t) + b_2\, a\, x_t(1-x_t) + \mu\, \xi_2
\end{aligned}
$$

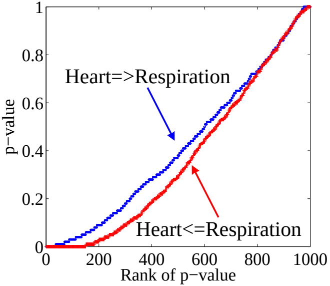
(Ancona et al. 2004) $a = 3.8$, $\mu = 0.01$, $\xi_{1,2} \propto \mathcal{N}(0,1)$



Real-world data: cardiorespiratory interaction

- Normally, Heart Rate ⇐ Respiration Force
- Sleep apnea affects the normal process of RSA (Respiratory Sinus Arrhythmia), disturbs the usual patterns: Heart ⇒ Respiration
  (also claimed by Schreiber 2000, Bhattacharya et al. 2003, Ancona et al. 2004)

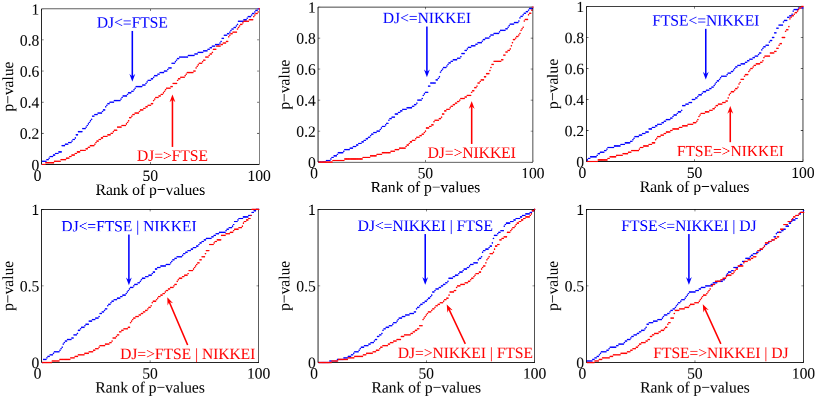




Data set B of Santa Fe Institute time series competition (Rigney et al. 1993)

Real-world data: co-movement of stock indexes I

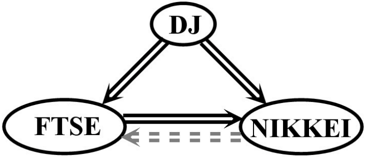
- Daily movements from April 1984 to January 2008
- Down Jones (DJ) industrial average index, Financial Times Stock Exchange (FTSE) 100, and NIKKEI 225



Real-world data: co-movement of stock indexes II

Test results of daily co-movements:

- DJ $\Rightarrow$ FTSE, and DJ $\Rightarrow$ FTSE | NIKKEI
- DJ $\Rightarrow$ NIKKEI, and DJ $\Rightarrow$ NIKKEI | FTSE
- FTSE $\Rightarrow$ NIKKEI, and FTSE $\Rightarrow$ NIKKEI | DJ
- FTSE $\Leftarrow$ NIKKEI, but FTSE $\not\Leftarrow$ NIKKEI | DJ



$\hookrightarrow$ "FTSE $\Leftarrow$ NIKKEI" might be spurious and mediated by DJ

Summary

- Subsampling-based kernel test of nonlinear Granger causality from time series data

Open issues:

- Connection to mutual information? (Gretton et al. 2005) or transfer entropy? (Schreiber 2000)

Thanks for your attention!