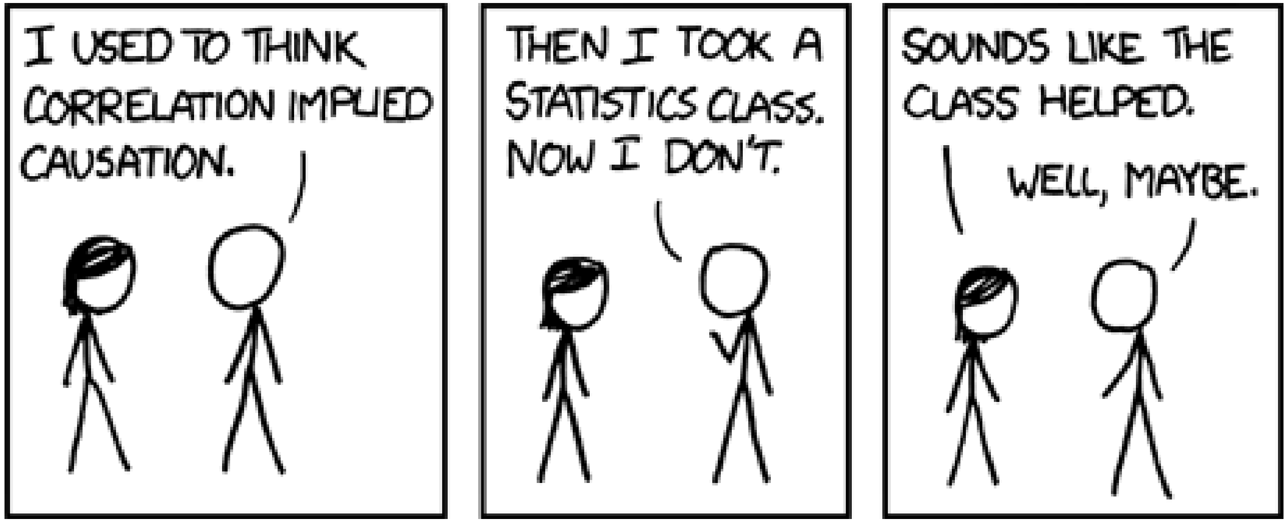
Causality with Non-Gaussian Time Series

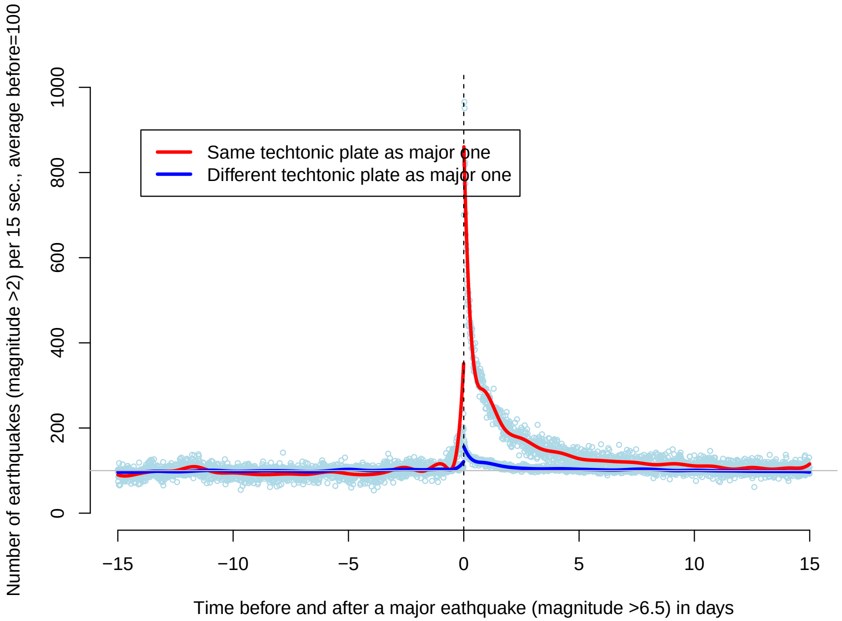
Arthur Charpentier (Université de Rennes 1 & UQàM)

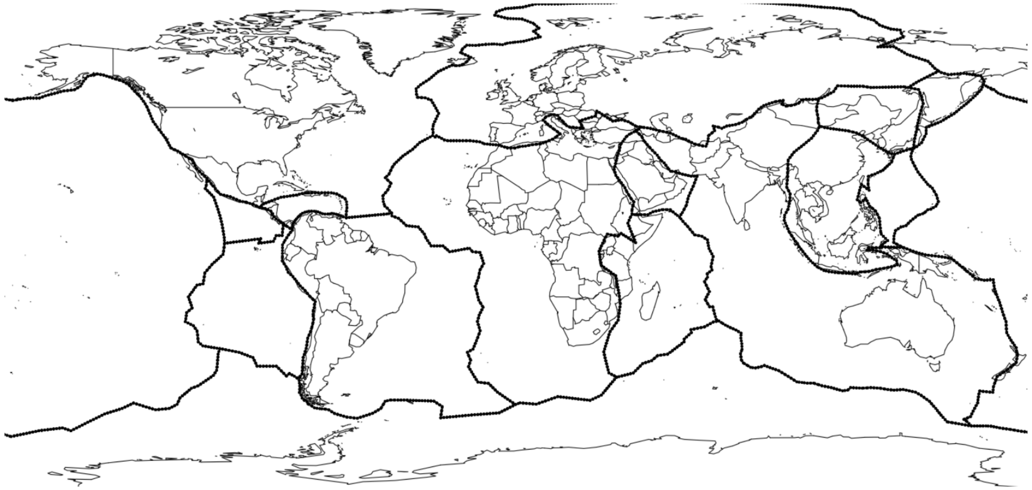


Université Paris 7 Diderot, May 2016.

http://freakonometrics.hypotheses.org

Motivation (Earthquakes)





see Boudreault & C. (2011) on contagion among tectonic plates

Motivation (Onsite vs. Online)

onsite *protestors*, *camped-out*, *arrests* and *injuries*

vs. online #indignados, #occupy and #vinegar on Twitter & Facebook

see Bastos, Mercea & C. (2015)

Multivariate Stationary Time Series

Definition A time series $(\boldsymbol{X}_t = (X_{1,t},\cdots,X_{d,t}))_{t\in\mathbb{Z}}$ with values in $\mathbb{R}^d$ is called a VAR(1) process if

$$\begin{cases} X_{1,t} = \phi_{1,1}X_{1,t-1} + \phi_{1,2}X_{2,t-1} + \cdots + \phi_{1,d}X_{d,t-1} + \varepsilon_{1,t} \\ X_{2,t} = \phi_{2,1}X_{1,t-1} + \phi_{2,2}X_{2,t-1} + \cdots + \phi_{2,d}X_{d,t-1} + \varepsilon_{2,t} \\ \cdots \\ X_{d,t} = \phi_{d,1}X_{1,t-1} + \phi_{d,2}X_{2,t-1} + \cdots + \phi_{d,d}X_{d,t-1} + \varepsilon_{d,t} \end{cases} \tag{1}$$

or equivalently

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{d,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,d} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,d} \\ \vdots & \vdots & & \vdots \\ \phi_{d,1} & \phi_{d,2} & \cdots & \phi_{d,d} \end{pmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{d,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{d,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}$$

Multivariate Stationary Time Series

For some real-valued $d \times d$ matrix Φ, and some i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ with values in $\mathbb{R}^d$.

Assume that $\boldsymbol{\varepsilon}_t$ is a Gaussian white noise $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, with density

$$f(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{(2\pi)^d |\det \boldsymbol{\Sigma}|}} \exp\left(-\frac{\boldsymbol{\varepsilon}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}}{2}\right), \quad \forall \boldsymbol{\varepsilon} \in \mathbb{R}^d.$$

Assume also that $\boldsymbol{\varepsilon}_t$ is independent of $\underline{\boldsymbol{X}}_{t-1} = \sigma(\{\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \cdots,\})$. : $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{Z}}$ is the innovation process.

Definition A time series $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is said to be (weakly) stationary if

- $\mathbb{E}(\boldsymbol{X}_t)$ is independent of t $(=: \boldsymbol{\mu})$
- $\text{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h})$ is independent of t $(=: \boldsymbol{\gamma}(h))$, called autocovariance matrix

Multivariate Stationary Time Series

Define the autocorrelation matrix,

$$\boldsymbol{\rho}(h) := \boldsymbol{\Delta}^{-1}\boldsymbol{\gamma}(h)\boldsymbol{\Delta}^{-1}, \text{ where } \boldsymbol{\Delta} := \sqrt{\text{diag}\,(\boldsymbol{\gamma}(0))}.$$

$(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ a stationary AR(1) time series, $\boldsymbol{X}_t = \boldsymbol{\Phi}\boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t$

Proposition $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is a stationary AR(1) time series if and only if the d eigenvalues of $\boldsymbol{\Phi}$ should have a norm lower than 1.

Proposition If $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is a stationary VAR(1) time series,

$$\boldsymbol{\rho}(h) = \boldsymbol{\Phi}^h, h \in \mathbb{N}.$$

Causality, in dimension 2

Two stationary time series $(X_t, Y_t)_{t\in\mathbb{Z}}$. Heuristics on independence,

$$f(x_t, y_t|\underline{X}_{t-1}, \underline{Y}_{t-1}) = f(x_t|\underline{X}_{t-1}) \cdot f(y_t|\underline{Y}_{t-1})$$

Write (with $\underline{X}$ for $\underline{X}_{t-1}$)

$$\underbrace{\frac{f(x_t, y_t|\underline{X}, \underline{Y})}{f(x_t|\underline{X}) \cdot f(y_t|\underline{Y})}}_{(X,Y)} = \underbrace{\frac{f(x_t|\underline{X}, \underline{Y})}{f(x_t|\underline{X})}}_{X\rightarrow Y} \cdot \underbrace{\frac{f(y_t|\underline{X}, \underline{Y})}{f(y_t|\underline{Y})}}_{X\leftarrow Y} \cdot \underbrace{\frac{f(x_t, y_t|\underline{X}, \underline{Y})}{f(x_t|\underline{X}, \underline{Y}) \cdot f(y_t|\underline{X}, \underline{Y})}}_{X\Leftrightarrow Y}$$

Gouriéroux, Monfort & Renault (1987) define the following Kullback-measures

$$C(X,Y) = \mathbb{E}\left[\log \frac{f(X_t, Y_t|\underline{X}, \underline{Y})}{f(X_t|\underline{X}) \cdot f(Y_t|\underline{Y})}\right]$$

Causality, in dimension 2

$$C(X \to Y) = \mathbb{E}\left[\log \frac{f(X_t|\underline{X},\underline{Y})}{f(X_t|\underline{X})}\right]$$

$$C(Y \to X) = \mathbb{E}\left[\log \frac{f(Y_t|\underline{X},\underline{Y})}{f(Y_t|\underline{Y})}\right]$$

$$C(X \Leftrightarrow Y) = \mathbb{E}\left[\log \frac{f(X_t,Y_t|\underline{X},\underline{Y})}{f(X_t|\underline{X},\underline{Y}) \cdot f(Y_t|\underline{X},\underline{Y})}\right]$$

so that $C(X,Y) = C(X \to Y) + C(X \leftarrow Y) + C(X \Leftrightarrow Y)$.

From Granger (1969)

(X) causes (Y) at time t if $\mathcal{L}(y_t|\underline{X}_{t-1},\underline{Y}_{t-1}) \neq \mathcal{L}(y_t|\underline{Y}_{t-1})$

(X) causes (Y) instantaneously at time t if $\mathcal{L}(y_t|\underline{X}_{t},\underline{Y}_{t-1}) \neq \mathcal{L}(y_t|\underline{X}_{t-1},\underline{Y}_{t-1})$

Causality, in dimension 2, for VAR(1) time series

$$\underbrace{\begin{pmatrix} X_t \\ Y_t \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{pmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} u_t \\ v_t \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with Var} \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$$

From Granger (1969) (see also Toda & Phillips (1994))

(X) causes (Y) at time t, $X \rightarrow Y$, if $\phi_{2,1} \neq 0$

(Y) causes (X) at time t, $Y \rightarrow X$, if $\phi_{1,2} \neq 0$

(X) causes (Y) instantaneously at time t, $X \Leftrightarrow X$, if $\sigma_{u,v} \neq 0$

Testing Causality, in dimension d

For lagged causality, we test

$$H_0 : \boldsymbol{\Phi} \in \mathcal{P} \text{ against } H_1 : \boldsymbol{\Phi} \notin \mathcal{P},$$

where $\mathcal{P}$ is a set of constrained shaped matrix, e.g. $\mathcal{P}$ is the set of $d \times d$ diagonal matrices for lagged independence, or a set of block triangular matrices for lagged causality.

Proposition Let $\widehat{\boldsymbol{\Phi}}$ denote the conditional maximum likelihood estimate of $\boldsymbol{\Phi}$ in the non-constrained MINAR(1) model, and $\widehat{\boldsymbol{\Phi}}^c$ denote the conditional maximum likelihood estimate of $\boldsymbol{\Phi}$ in the constrained model, then under suitable conditions,

$$2[\log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{\Phi}}|\boldsymbol{X}_0) - \log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{\Phi}}^c|\boldsymbol{X}_0)] \overset{\mathcal{L}}{\to} \chi^2(d^2 - \dim(\mathcal{P})), \text{ as } T \to \infty, \text{ under } H_0.$$

Example Testing $(X_{1,t}) \leftarrow (X_{2,t})$ is testing whether $\phi_{1,2} = 0$, or not.

Modeling Counts Processes

Steutel & van Harn (1979) defined a thinning operator as follows

Definition Define operator $\circ$ as

$$p \circ N = \sum_{i=1}^{N} Y_i = Y_1 + \cdots + Y_N \text{ if } N \neq 0, \text{ and } 0 \text{ otherwise},$$

where N is a random variable with values in $\mathbb{N}$, $p \in [0,1]$, and $Y_1, Y_2, \cdots$ are i.i.d. Bernoulli variables, independent of N, with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$. Thus $p \circ N$ is a compound sum of i.i.d. Bernoulli variables.

Hence, given N, $p \circ N$ has a binomial distribution $\mathcal{B}(N, p)$.

Note that $p \circ (q \circ \boldsymbol{N}) \stackrel{\mathcal{L}}{=} [pq] \circ \boldsymbol{N}$ for all $p, q \in [0,1]$.

Further

$$\mathbb{E}\left(p \circ N\right) = p\mathbb{E}(N) \text{ and } \mathrm{Var}\left(p \circ N\right) = p^2 \mathrm{Var}(N) + p(1-p)\mathbb{E}(N).$$

(Poisson) Integer AutoRegressive processes $INAR(1)$

Based on that thinning operator, Al-Osh & Alzaid (1987) and McKenzie (1985) defined the integer autoregressive process of order 1:

Definition A time series $(X_t)_{t\in\mathbb{N}}$ with values in $\mathbb{R}$ is called an INAR(1) process if

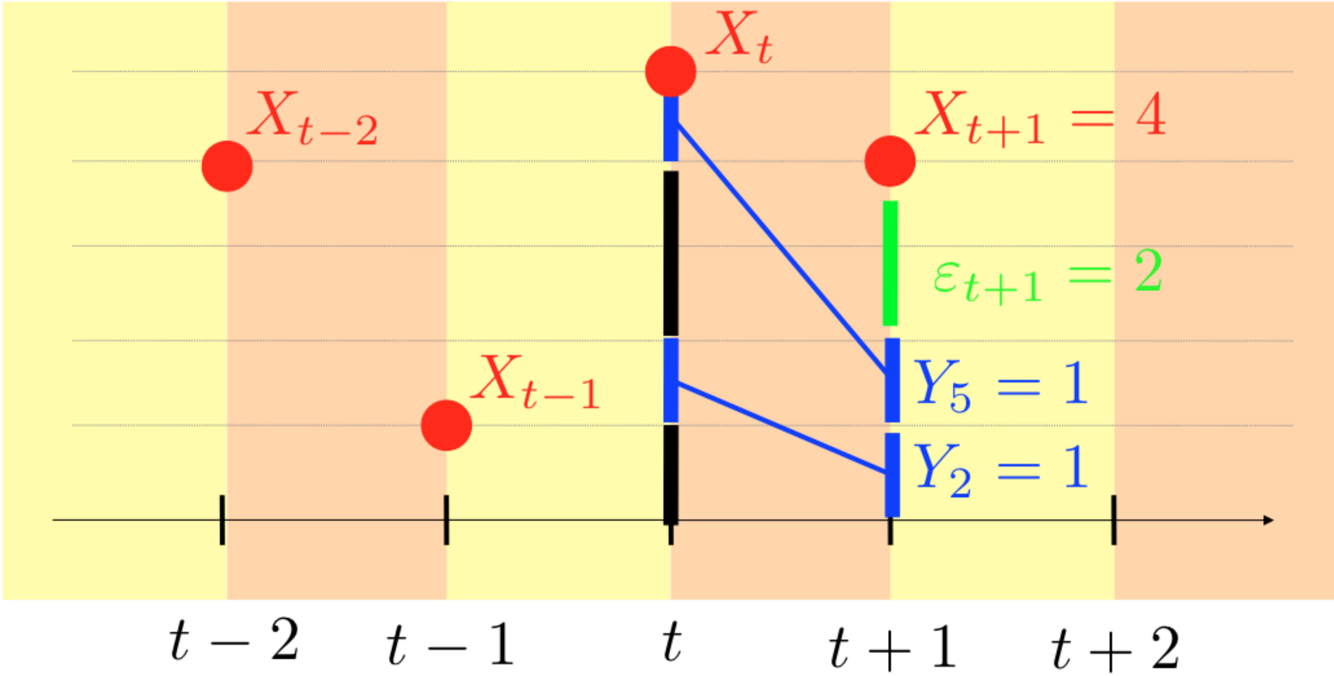
$$X_t = p \circ X_{t-1} + \varepsilon_t, \tag{2}$$

where (ε_t) is a sequence of i.i.d. integer valued random variables, i.e.

$$X_t = \sum_{i=1}^{X_{t-1}} Y_i + \varepsilon_t, \text{ where } Y_i's \text{ are } i.i.d.\ \mathcal{B}(p).$$

Such a process can be related to Galton-Watson processes.

$INAR(1)$ & Galton-Watson



$$X_{t+1}=\sum_{i=1}^{X_t} Y_i+\varepsilon_{t+1},\text{ where } Y_i's \text{ are } i.i.d.\ \mathcal{B}(p)$$

Proposition $\mathbb{E}(X_t) = \dfrac{\mathbb{E}(\varepsilon_t)}{1-p}$, $\text{Var}(X_t) = \gamma(0) = \dfrac{p\mathbb{E}(\varepsilon_t) + \text{Var}(\varepsilon_t)}{1-p^2}$ and

$$\gamma(h) = \text{cov}(X_t, X_{t-h}) = p^h.$$

It is common to assume that ε_t are independent variables, with a Poisson distribution $\mathcal{P}(\lambda)$, with probability function

$$\mathbb{P}(\varepsilon_t = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k \in \mathbb{N}.$$

Proposition If (ε_t) are Poisson random variables, then (X_t) will also be a sequence of Poisson random variables.

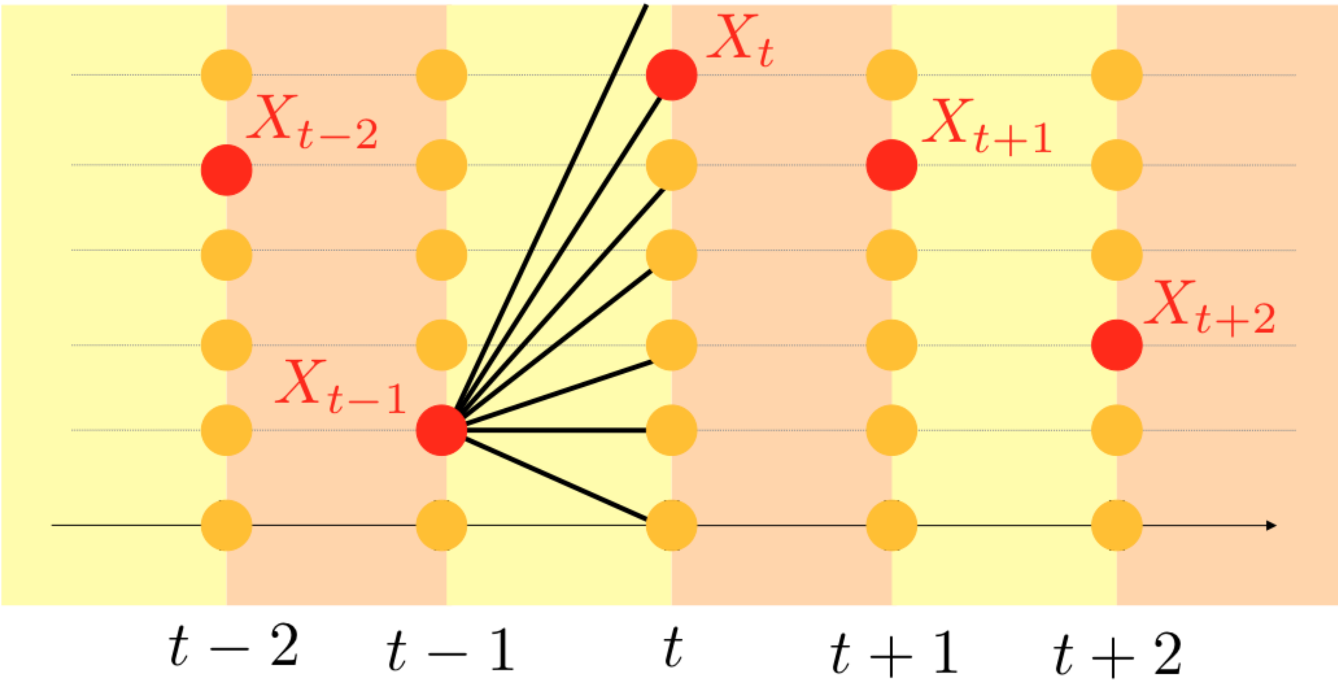
Note that we assume also that ε_t is independent of $\underline{X}_{t-1}$, i.e. past observations $X_0, X_1, \cdots, X_{t-1}$. Thus, $(\varepsilon_t)_{t\in\mathbb{N}}$ is called the innovation process.

Proposition $(X_t)_{t\in\mathbb{N}}$ is a stationary $INAR(1)$ time series if and only if $p \in [0, 1)$.

Proposition If $(X_t)_{t\in\mathbb{N}}$ is a stationary $INAR(1)$ time series, $(X_t)_{t\in\mathbb{N}}$ is an homogeneous Markov chain.

Markov Property of $INAR(1)$ Time Series

$$\pi(x_t, x_{t-1}) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) = \sum_{k=0}^{x_t} \underbrace{\mathbb{P}\left(\sum_{i=1}^{x_{t-1}} Y_i = x_t - k\right)}_{\text{Binomial}} \cdot \underbrace{\mathbb{P}(\varepsilon = k)}_{\text{Poisson}}.$$



Inference of $INAR(1)$ Processes

Consider a Poisson $INAR(1)$ process, then the likelihood is

$$\mathcal{L}(p,\lambda;X_0,\underline{\boldsymbol{X}})=\left[\prod_{t=1}^{n}f_t(X_t)\right]\cdot\frac{\lambda^{X_0}}{(1-p)^{X_0}X_0!}\exp\left(-\frac{\lambda}{1-p}\right)$$

where

$$f_t(y)=\exp(-\lambda)\sum_{i=0}^{\min\{X_t,X_{t-1}\}}\frac{\lambda^{y-i}}{(y-i)!}\binom{Y_{t-1}}{i}p^i(1-p)^{Y_{t-1}-y},\text{ for }t=1,\cdots,n.$$

Maximum likelihood estimators are

$$(\widehat{p},\widehat{\lambda})\in\text{argmax}\left\{\log\mathcal{L}(p,\lambda;(X_0,\boldsymbol{\mathcal{X}}))\right\}$$

Multivariate Integer Autoregressive processes $MINAR(1)$

Let $\boldsymbol{X_t} := (X_{1,t}, \cdots, X_{d,t})$, denote a multivariate vector of counts.

Definition Let $\boldsymbol{P} := [p_{i,j}]$ be a $d \times d$ matrix with entries in $[0, 1]$. If $\boldsymbol{X} = (X_1, \cdots, X_d)$ is a random vector with values in $\mathbb{N}^d$, then $\boldsymbol{P} \circ \boldsymbol{X}$ is a d-dimensional random vector, with i-th component

$$[\boldsymbol{P} \circ \boldsymbol{X}]_i = \sum_{j=1}^{d} p_{i,j} \circ X_j,$$

for all $i = 1, \cdots, d$, where all counting variates Y in $p_{i,j} \circ X_j$'s are assumed to be independent.

Note that $\boldsymbol{P} \circ (\boldsymbol{Q} \circ \boldsymbol{X}) \overset{\mathcal{L}}{=} [\boldsymbol{P}\boldsymbol{Q}] \circ \boldsymbol{X}$.

Further, $\mathbb{E}\left(\boldsymbol{P} \circ \boldsymbol{X}\right) = \boldsymbol{P}\mathbb{E}(\boldsymbol{X})$, and

$$\mathbb{E}\left((\boldsymbol{P} \circ \boldsymbol{X})(\boldsymbol{P} \circ \boldsymbol{X})^{\mathsf{T}}\right) = \boldsymbol{P}\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}})\boldsymbol{P}^{\mathsf{T}} + \Delta,$$

with $\Delta := \text{diag}(\boldsymbol{V}\mathbb{E}(\boldsymbol{X}))$ where $\boldsymbol{V}$ is the $d \times d$ matrix with entries $p_{i,j}(1 - p_{i,j})$.

Multivariate Integer Autoregressive processes $MINAR(1)$

Definition A time series $(\boldsymbol{X}_t)$ with values in $\mathbb{N}^d$ is called a d-variate MINAR(1) process if

$$\boldsymbol{X}_t = \boldsymbol{P} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{3}$$

for all t, for some $d \times d$ matrix $\boldsymbol{P}$ with entries in $[0,1]$, and some i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ with values in $\mathbb{N}^d$.

$(\boldsymbol{X}_t)$ is a Markov chain with states in $\mathbb{N}^d$ with transition probabilities

$$\pi(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) = \mathbb{P}(\boldsymbol{X}_t = \boldsymbol{x}_t | \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) \tag{4}$$

satisfying

$$\pi(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) = \sum_{\boldsymbol{k}=0}^{\boldsymbol{x}_t} \mathbb{P}(\boldsymbol{P} \circ \boldsymbol{x}_{t-1} = \boldsymbol{x}_t - \boldsymbol{k}) \cdot \mathbb{P}(\boldsymbol{\varepsilon} = \boldsymbol{k}).$$

Inference for $MINAR(1)$

Proposition Let $(\boldsymbol{X}_t)$ be a d-variate MINAR(1) process satisfying stationary conditions, as well as technical assumptions (called C1-C6 in Franke & Subba Rao (1993)), then the conditional maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\boldsymbol{P}, \boldsymbol{\Lambda})$ is asymptotically normal,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma^{-1}(\boldsymbol{\theta})), \text{ as } n \to \infty.$$

Further,

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\theta}} | \boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \boldsymbol{\theta} | \boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 + \dim(\boldsymbol{\lambda})), \text{ as } n \to \infty.$$

Granger causality with $BINAR(1)$

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with Var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

1. (X_1) and (X_2) are instantaneously related if $\boldsymbol{\varepsilon}$ is a noncorrelated noise,

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with } \mathrm{Var}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \star \\ \star & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

2. (X_1) and (X_2) are independent, $(X_1)\perp(X_2)$ if $\boldsymbol{P}$ is diagonal, i.e. $p_{1,2}=p_{2,1}=0$, and ε_1 and ε_2 are independent,

$$\underbrace{\begin{pmatrix}X_{1,t}\\X_{2,t}\end{pmatrix}}_{\boldsymbol{X}_t}=\underbrace{\begin{pmatrix}p_{1,1}&0\\0&p_{2,2}\end{pmatrix}}_{\boldsymbol{P}}\circ\underbrace{\begin{pmatrix}X_{1,t-1}\\X_{2,t-1}\end{pmatrix}}_{\boldsymbol{X}_{t-1}}+\underbrace{\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}}_{\boldsymbol{\varepsilon}_t},\text{ with Var}\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}=\begin{pmatrix}\lambda_1&0\\0&\lambda_2\end{pmatrix}$$

Granger causality with $BINAR(1)$

3. (N_1) causes (N_2) but (N_2) does not cause (X_1), $(X_1)\rightarrow(X_2)$, if $\boldsymbol{P}$ is a lower triangle matrix, i.e. $p_{2,1}\neq 0$ while $p_{1,2}=0$,

$$\underbrace{\begin{pmatrix}X_{1,t}\\X_{2,t}\end{pmatrix}}_{\boldsymbol{X}_t}=\underbrace{\begin{pmatrix}p_{1,1}&0\\\star&p_{2,2}\end{pmatrix}}_{\boldsymbol{P}}\circ\underbrace{\begin{pmatrix}X_{1,t-1}\\X_{2,t-1}\end{pmatrix}}_{\boldsymbol{X}_{t-1}}+\underbrace{\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}}_{\boldsymbol{\varepsilon}_t},\text{ with Var}\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}=\begin{pmatrix}\lambda_1&\varphi\\\varphi&\lambda_2\end{pmatrix}$$

Granger causality with $BINAR(1)$

4. (N_2) causes (N_1) but $(N_{1,t})$ does not cause (N_2), $(N_1)\leftarrow(N_{2,t})$, if $\boldsymbol{P}$ is a upper triangle matrix, i.e. $p_{1,2}\neq 0$ while $p_{2,1}=0$,

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & \star \\ 0 & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with Var}\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

5. (N_1) causes (N_2) and conversely, i.e. a feedback effect $(N_1)\leftrightarrow(N_2)$, if $\boldsymbol{P}$ is a full matrix, i.e. $p_{1,2}, p_{2,1} \neq 0$

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & \star \\ \star & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with Var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Bivariate Poisson $BINAR(1)$

A classical distribution for $\boldsymbol{\varepsilon}_t$ is the bivariate Poisson distribution, with one common shock, i.e.

$$\begin{cases} \varepsilon_{1,t} = M_{1,t} + M_{0,t} \\ \varepsilon_{2,t} = M_{2,t} + M_{0,t} \end{cases}$$

where $M_{1,t}$, $M_{2,t}$ and $M_{0,t}$ are independent Poisson variates, with parameters $\lambda_1 - \varphi$, $\lambda_2 - \varphi$ and φ, respectively. In that case, $\boldsymbol{\varepsilon}_t := (\varepsilon_{1,t}, \varepsilon_{2,t})$ has joint probability function

$$e^{-[\lambda_1+\lambda_2-\varphi]} \frac{(\lambda_1-\varphi)^{k_1}}{k_1!} \frac{(\lambda_2-\varphi)^{k_2}}{k_2!} \sum_{i=0}^{\min\{k_1,k_2\}} \binom{k_1}{i}\binom{k_2}{i} i! \left(\frac{\varphi}{[\lambda_1-\varphi][\lambda_2-\varphi]}\right)$$

with $\lambda_1, \lambda_2 > 0$, $\varphi \in [0, \min\{\lambda_1, \lambda_2\}]$.

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Bivariate Poisson $BINAR(1)$ and Granger causality

For instantaneous causality, we test

$$H_0 : \varphi = 0 \text{ against } H_1 : \varphi \neq 0$$

Proposition Let $\widehat{\boldsymbol{\lambda}}$ denote the conditional maximum likelihood estimate of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \varphi)$ in the non-constrained MINAR(1) model, and $\boldsymbol{\lambda}^\perp$ denote the conditional maximum likelihood estimate of $\boldsymbol{\lambda}^\perp = (\lambda_1, \lambda_2, 0)$ in the constrained model (when innovation has independent margins), then under suitable conditions,

$$2[\log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{\lambda}}|\boldsymbol{X}_0) - \log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{\lambda}}^\perp|\boldsymbol{X}_0)] \xrightarrow{\mathcal{L}} \chi^2(1), \text{ as } n \to \infty, \text{ under } H_0.$$

Bivariate Poisson $BINAR(1)$ and Granger causality

For lagged causality, we test

$$H_0 : \boldsymbol{P} \in \mathcal{P} \text{ against } H_1 : \boldsymbol{P} \notin \mathcal{P},$$

where $\mathcal{P}$ is a set of constrained shaped matrix, e.g. $\mathcal{P}$ is the set of $d \times d$ diagonal matrices for lagged independence, or a set of block triangular matrices for lagged causality.

Proposition Let $\widehat{\boldsymbol{P}}$ denote the conditional maximum likelihood estimate of $\boldsymbol{P}$ in the non-constrained MINAR(1) model, and $\widehat{\boldsymbol{P}}^c$ denote the conditional maximum likelihood estimate of $\boldsymbol{P}$ in the constrained model, then under suitable conditions,

$$2[\log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{P}}|\boldsymbol{X}_0) - \log \mathcal{L}(\underline{\boldsymbol{X}}, \widehat{\boldsymbol{P}}^c|\boldsymbol{X}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 - \dim(\mathcal{P})), \text{ as } n \to \infty, \text{ under } H_0.$$

Example Testing $(X_{1,t}) \leftarrow (X_{2,t})$ is testing whether $p_{1,2} = 0$, or not.

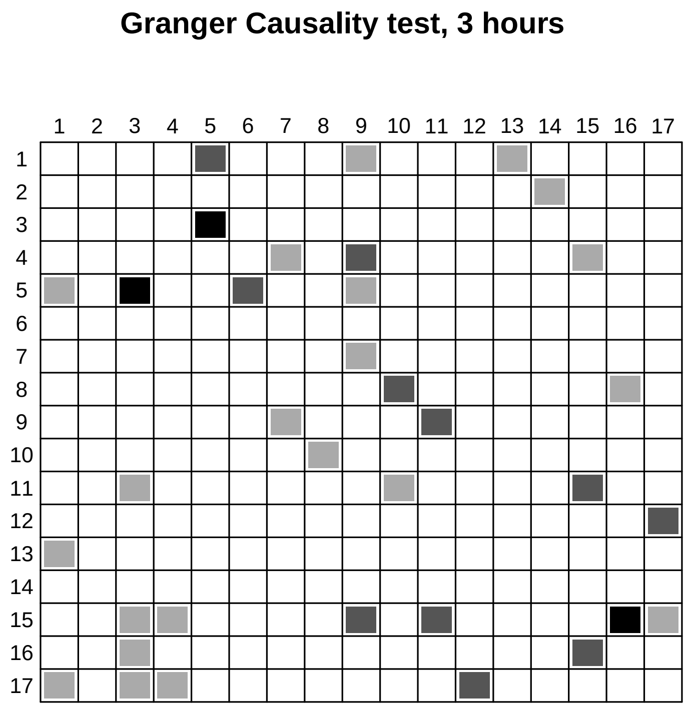
Autocorrelation of $MINAR(1)$ processes

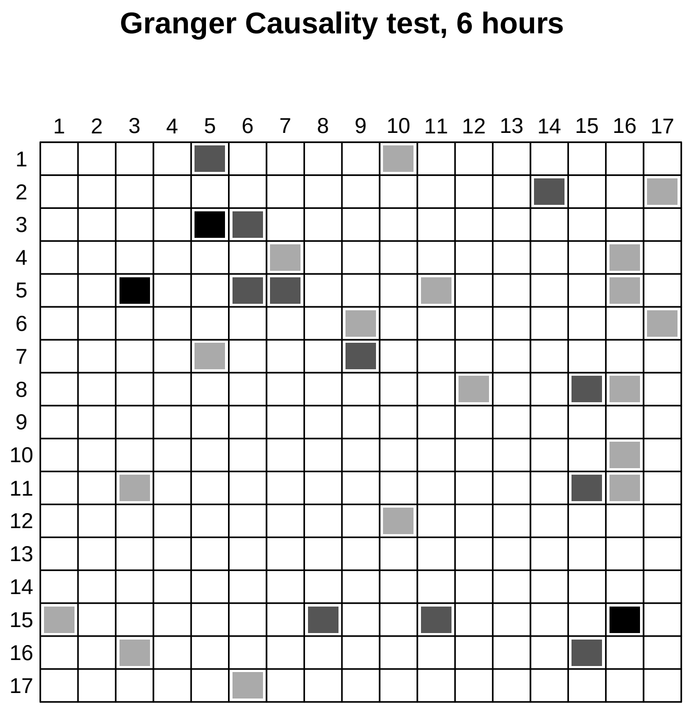
Proposition Consider a MINAR(1) process with representation $\boldsymbol{X}_t = \boldsymbol{P} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t$, where $(\boldsymbol{\varepsilon}_t)$ is the innovation process, with $\boldsymbol{\lambda} := \mathbb{E}(\boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\Lambda} := \text{Var}(\boldsymbol{\varepsilon}_t)$. Let $\boldsymbol{\mu} := \mathbb{E}(\boldsymbol{X}_t)$ and $\boldsymbol{\gamma}(h) := \text{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h})$. Then $\boldsymbol{\mu} = [\mathbb{I} - \boldsymbol{P}]^{-1}\boldsymbol{\lambda}$ and for all $h \in \mathbb{Z}$, $\boldsymbol{\gamma}(h) = \boldsymbol{P}^h\boldsymbol{\gamma}(0)$ with $\boldsymbol{\gamma}(0)$ solution of $\boldsymbol{\gamma}(0) = \boldsymbol{P}\boldsymbol{\gamma}(0)\boldsymbol{P}^\mathsf{T} + (\boldsymbol{\Delta} + \boldsymbol{\Lambda})$.

See Boudreault & C. (2011) for additional properties

Granger causality $X_1 \rightarrow X_2$ or $X_1 \leftarrow X_2$

1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

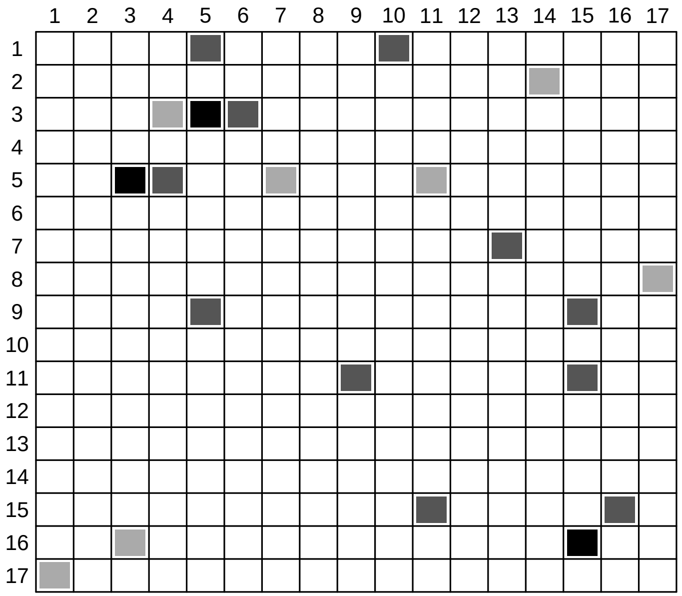




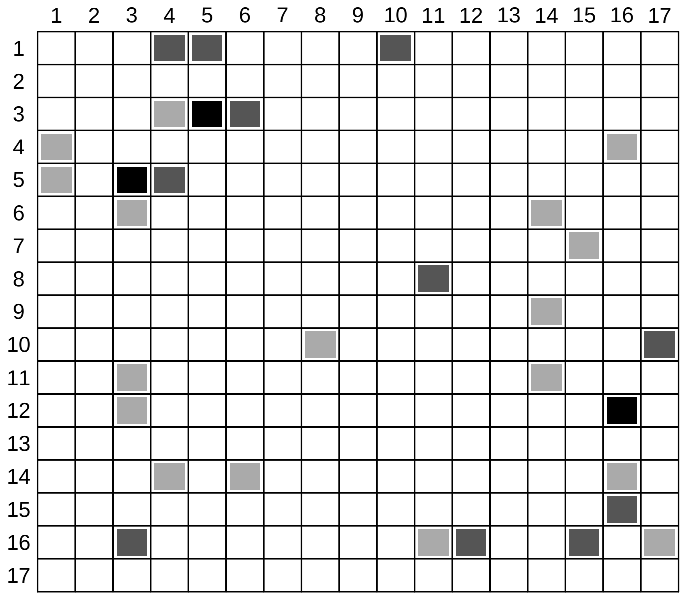
Granger causality $X_1 \rightarrow X_2$ or $X_1 \leftarrow X_2$

1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

Granger Causality test, 12 hours



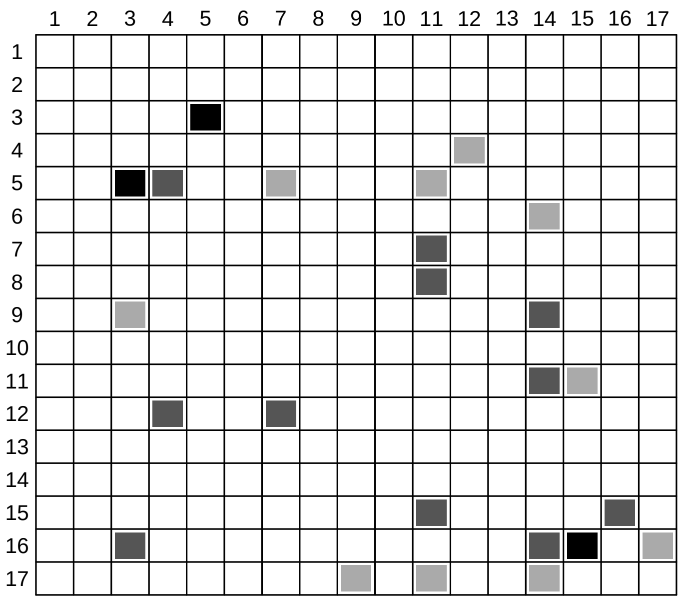
Granger Causality test, 24 hours



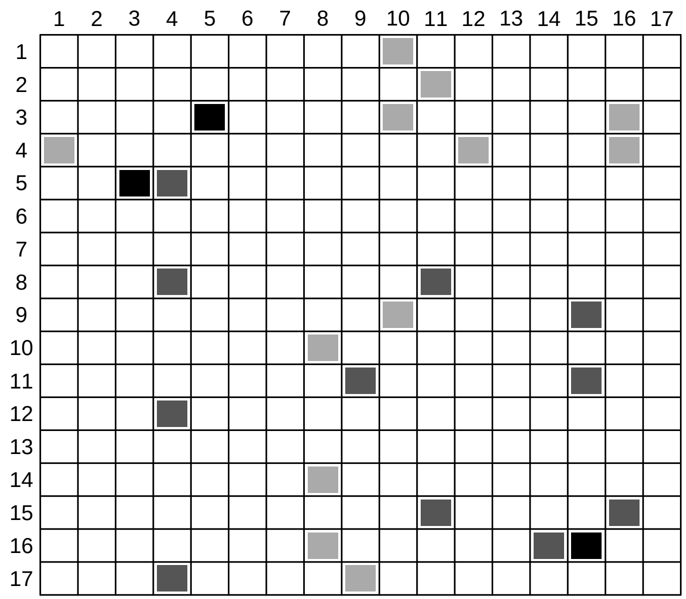
Granger causality $X_1 \to X_2$ or $X_1 \leftarrow X_2$

1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

Granger Causality test, 36 hours



Granger Causality test, 48 hours



Using Ranks for Time Series

Haugh (1976) suggested to use ranks to test for independence.

Set R_t denote the rank of X_t within $\{X_1, \cdots, X_T\}$, and set

$$U_t = \frac{R_t}{T} = \frac{1}{T}\sum_{s=1}^{T}\mathbf{1}_{X_t \leq X_s} = \widehat{F}_X(X_t)$$

and similarly

$$V_t = \frac{S_t}{T} = \frac{1}{T}\sum_{s=1}^{T}\mathbf{1}_{Y_t \leq Y_s} = \widehat{F}_Y(Y_t)$$

See also Dufour(1981) for rank tests for serial dependence.

Causality, in dimension 2

From Taamouti, Bouezmarni & El Ghouch (2014), consider some copula based causality approach:

$$C(X \to Y) = \mathbb{E}\left[\log \frac{f(X_t|\underline{X},\underline{Y})}{f(X_t|\underline{X})}\right]$$

can be written, for Markov 1 processes

$$C(X \to Y) = \mathbb{E}\left[\log \frac{f(X_t|X_{t-1},Y_{t-1})}{f(X_t|X_{t-1})}\right] = \mathbb{E}\left[\log \frac{f(X_t,X_{t-1},Y_{t-1})\cdot f(X_{t-1})}{f(X_t,X_{t-1})\cdot f(X_{t-1},Y_{t-1})}\right]$$

i.e.

$$C(X \to Y) = \mathbb{E}\left[\log \frac{c(F_X(X_t),F_X(X_{t-1}),F_Y(Y_{t-1}))}{c(F_X(X_t),F_X(X_{t-1}))\cdot c(F_X(X_{t-1}),F_Y(Y_{t-1}))}\right]$$

Using a Probit-type Transformation

Following Geenens, C. & Paindaveine (2014), consider some Probit-type transformation, for stationary time series

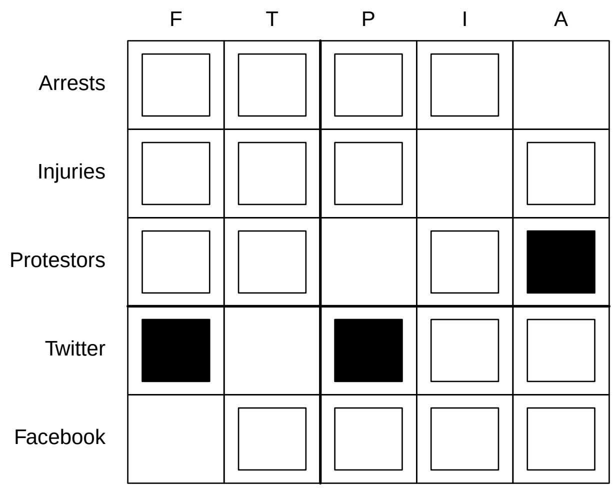
$$\widetilde{X}_t = \Phi^{-1}(U_t) = \Phi^{-1}(\widehat{F}_X(X_t))$$

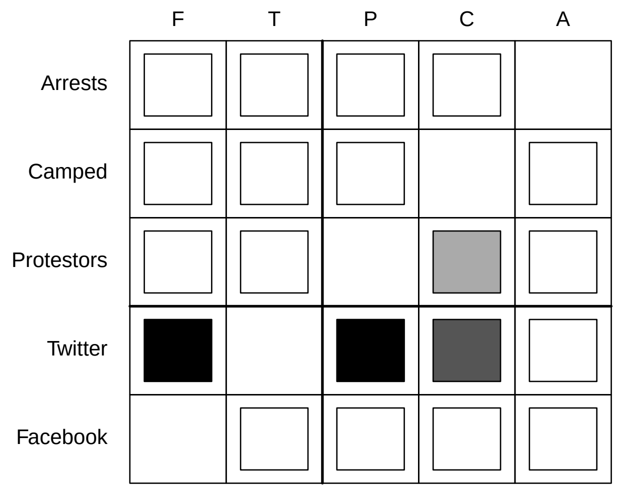
$$\widetilde{Y}_t = \Phi^{-1}(V_t) = \Phi^{-1}(\widehat{F}_Y(Y_t))$$

Application in Bastos, Mercea & C. (2015)

Online vs. Onsite Causality

For #occupy and #indignados





Application in Bastos, Mercea & C. (2015)