Time series causality inference using the Phase Slope Index.

Florin Popescu
Guido Nolte
Fraunhofer Institute FIRST, Berlin

Fraunhofer
FIRST

Introduction

- Linear time series analysis techniques can be useful in analyzing data that is actually generated by **nonlinear** stochastic processes (i.e. in the real world).
- Linear time series analysis can be conducted in the time domain (e.g. autoregressive models) or in the frequency domain (e.g. discrete Fourier transform, coherency among spectra) – theoretically both approaches are equivalent but numerically they are not. Causal estimation in time domain (AR): Granger 1973, Kaminski Blinowska 1991, Schreiber 2000, Rosenblum & Pikovsky 2001. Frequency domain method: Phase Slope Index (Nolte et al. 2008, Nolte et al. 2009) . Connection: partially directed coherence (Baccala & Sameshima 1998, 2001).
- Separating **correlation** from **causation** is hard, **even if** the data is time-labeled. There can be correlations among non-interacting time-series variables.

Outline

- Overview of different types of data generating processes (DGPs), which are stochastic generative models of time series
- Highlight causality assessment challenges in neuroscience and economics.
- AR estimation challenges for covariate innovations processes (needed for GC).
- PSI - Phase Slope Index
- PSI and AR results for bi-variate simulations available on Causality Workbench.
- Structural causality estimation in multivariate time series.

DGP: Data Generating Process

Data Generating Process

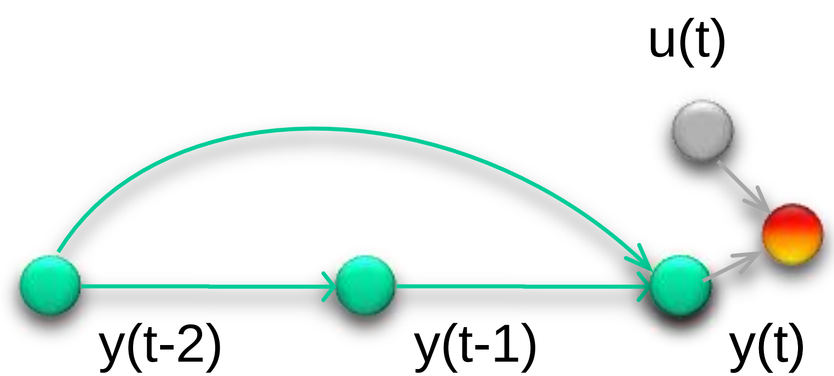
$$y(t) = q(y(t-1), y(t-2), ..y(t-p), u(t))$$

○ DGPs are abstractions of real-world dynamic processes which generate data: not necessarily are they regressive, recursive or stochastic, but are more powerful when they are.

○ They can be **inferred** from data directly or by **bottom-up** modeling of the underlying physical /social processes (in neuroscience, economics very hard)

DGP Symbolic representation

$$u(t) \perp u(t-1)$$



$p = 2$

$\Rightarrow_S$

Stochastic DGP

Data Generating Process

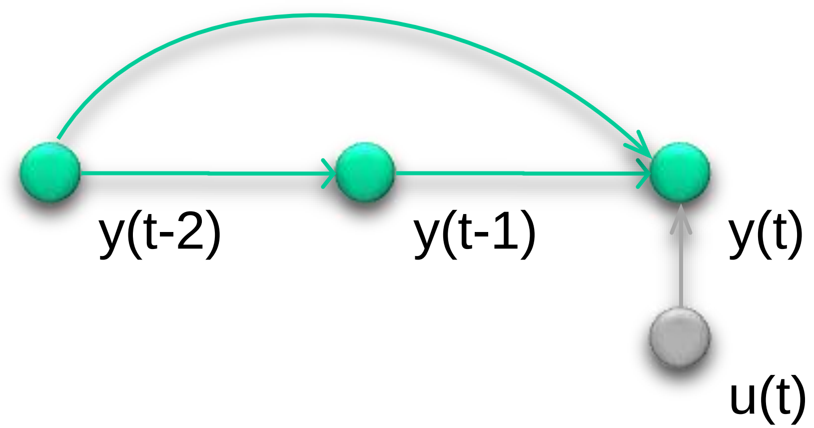
$$y(t) = q(y(t-1), y(t-2), ..y(t-p), u(t))$$

○ If the DGP is stochastic and noise in an input it is generally called **innovations process** and it is independently distributed if $p(u(t)) \perp t$ it is independently distributed.

○ If, also $q(y, u) \perp t$ then the system is **stationary**.

$$u(t) \perp y(t)$$

DGP Symbolic representation



DGP equivalence

Equivalence:

2 DGPs are **output** equivalent if, for all t:

$$Y_{1..t} = Q_1(t, u(t)) = Q_2(t, u(t))$$

$$Q_1 \equiv_* \ Q_2$$

DGPs are **stochastically** equivalent if, for all t:

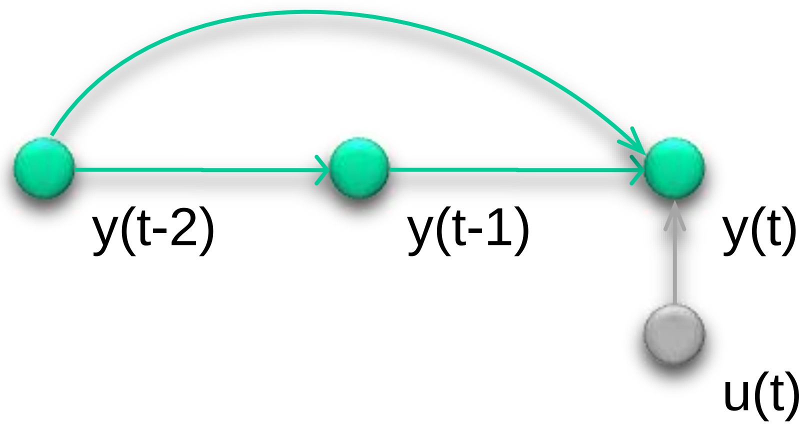
$$p(Q_1(t)) = p(Q_2(t))$$

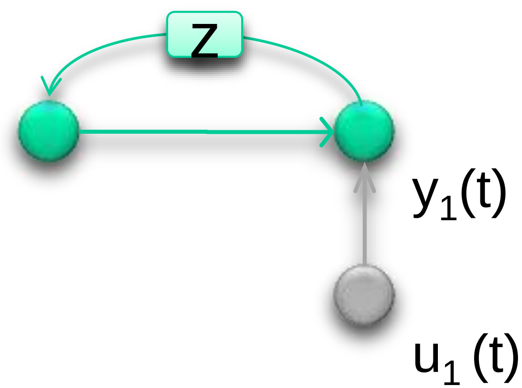
$$Q_1 \equiv_p \ Q_2$$

$$y_1(t) = \{y(t-1), y(t-2), ..y(t-p)\}$$

Canonical representation (non-unique)

DGP Symbolic representation





DGP variations

Potential DGP ‘upgrades’

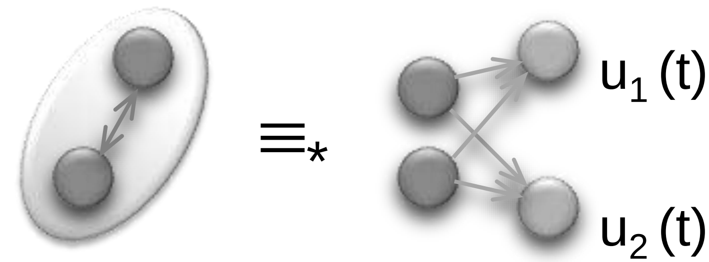
○ **covariate or mixed innovations**

$u_{1,2} = \mathcal{N}(0, \Sigma)$
is stochastically equivalent to $u_{1,2} = Bw_{1,2}$
where $w_{1,2} = \mathcal{N}(0, I)$ and $B'B = \Sigma$
for all C whuch satisfy $C'C = B'B$

○ **endogenous/**exogenous **inputs**

○ **cointegration**

DGP Symbolic representation



DGP variations

Data Generating Process

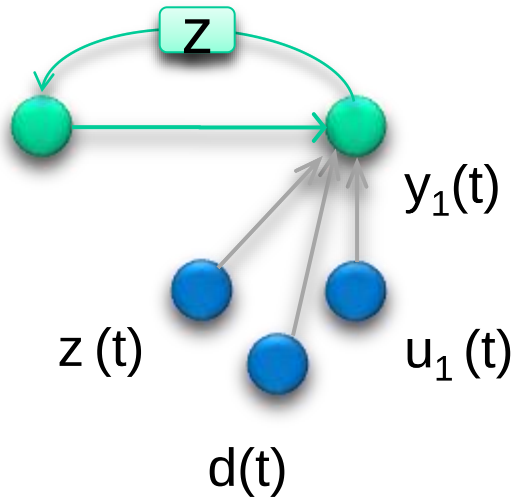
○ **covariate or mixed innovations**

○ **endogenous/**exogenous **inputs**

some inputs are stochastic ***but observable***, or non-stochastic, or excluded from potential effects

○ **co-integration**

DGP Symbolic representation



DGP variations

Potential DGP 'upgrades'

○ **covariate or mixed innovations**

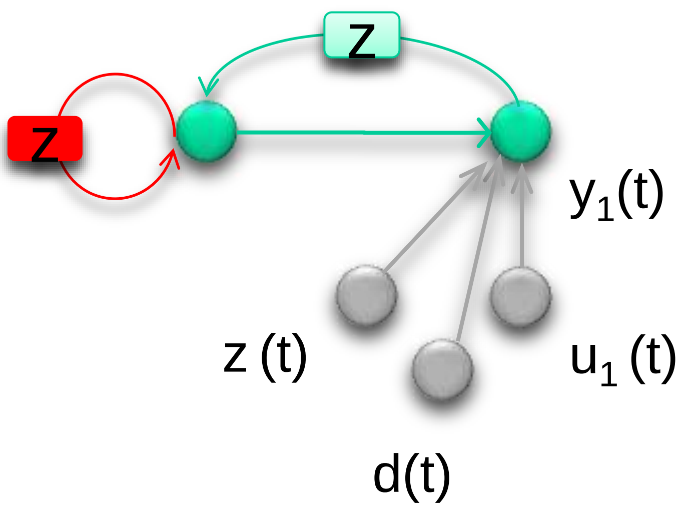
○ **endogenous/**exogenous **inputs**

some inputs are stochastic ***but observable***, or simply non-stochastic

○ **co-integration**

Some states are simple dynamic transformations of i.i.d processes -this can be taken into account

DGP Symbolic representation



G - Causality

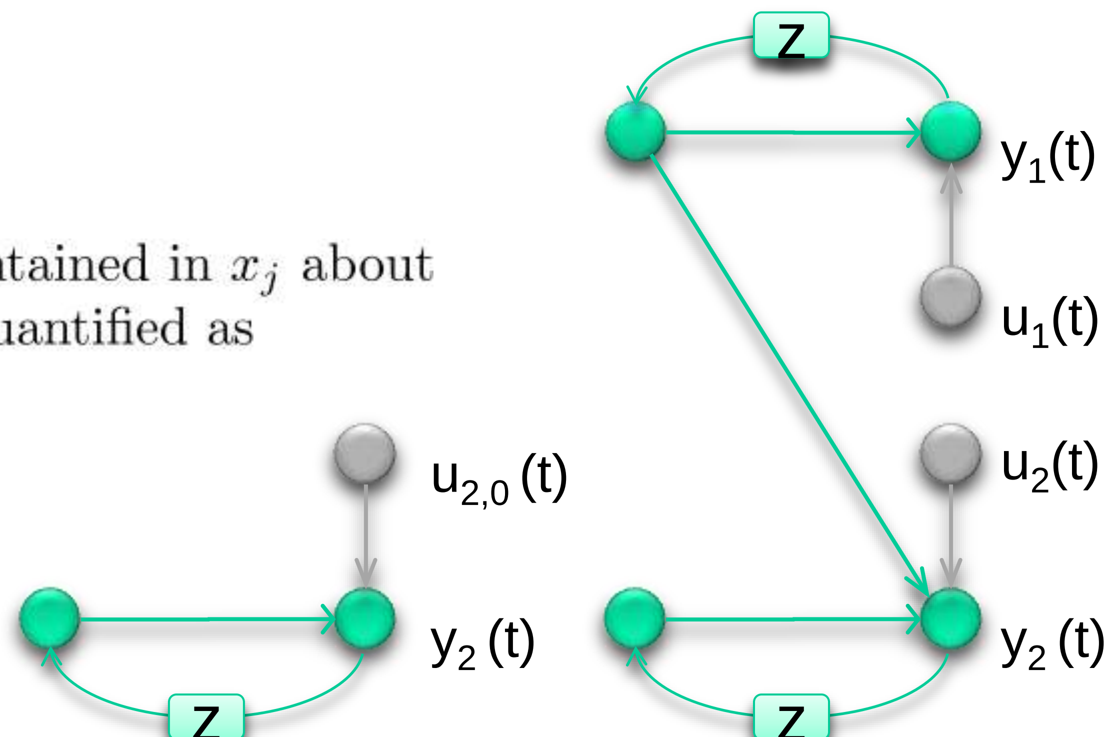
Granger causality inference requires derivation of a predictive model (can of worms…)

The additional information contained in x_j about the future of x_i for $j \neq i$ can be quantified as

$$\Gamma_{j\rightarrow i} = \log\left(\frac{Var(u_{i,0})}{Var(u_i)}\right)$$

$\Gamma_{j\rightarrow i} > 0$ one says that channel j 'G-causes' channel i.

DGP Symbolic representation



G-causality

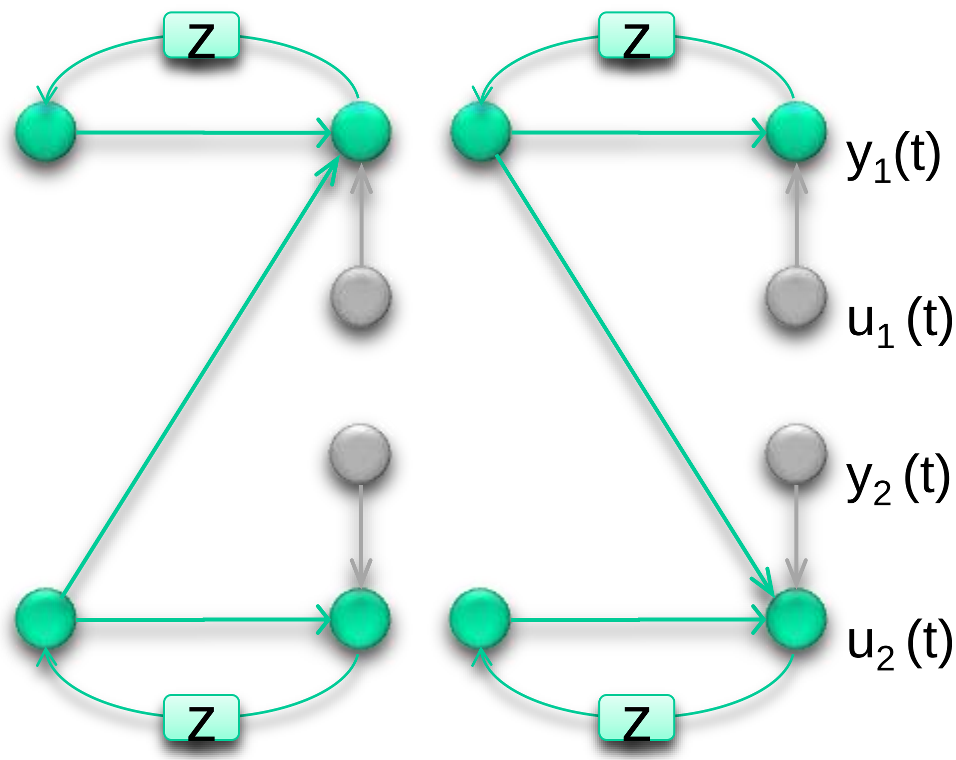
○ G-causality is inferred by comparing conditional entropy in competing structural models

We can establish a statistic:
$\tilde{G} = \Gamma_{1\to 2} - \Gamma_{2\to 1}$
to be normalized by a jacknife-derived std
$G = \frac{\tilde{G}}{std(\tilde{G})}$
$abs(G) > 2$ can be viewed as significant.

$1 \to 2$ $1 \leftarrow 2$

$1 \perp 2$ $1 \leftrightarrow 2$

DGP Symbolic representation



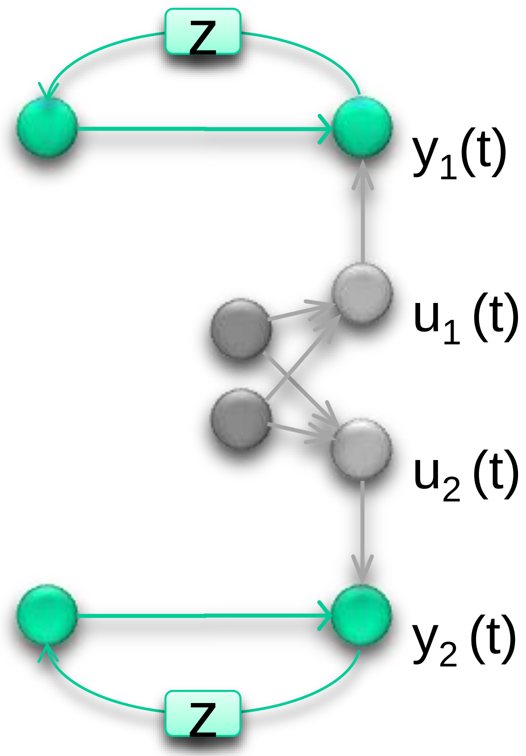
Structural / G - Causality

Covariate innovations?

○ In many instances it is reasonable to assume that the innovations process is covariate. For example: yearly weather variability and historical shocks on aggregate indicators.

○ Also possible is that other unobservable factors actually provide root causes for correlations among innovations processes.

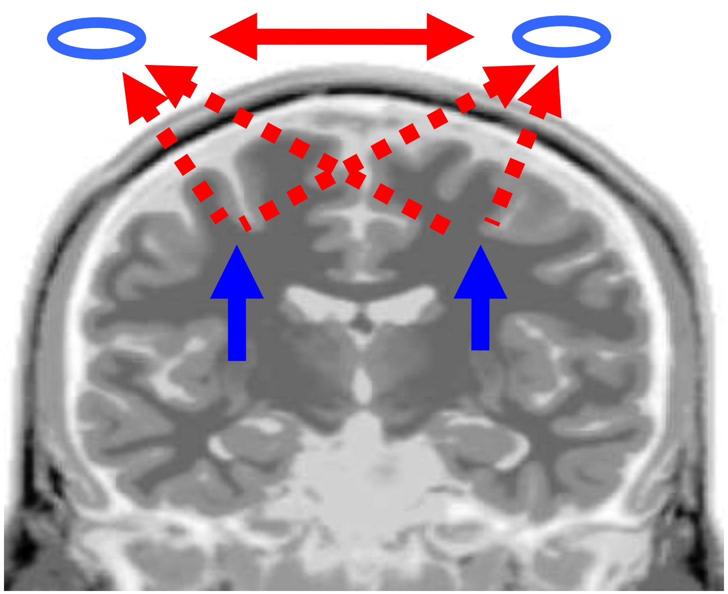
DGP Symbolic representation



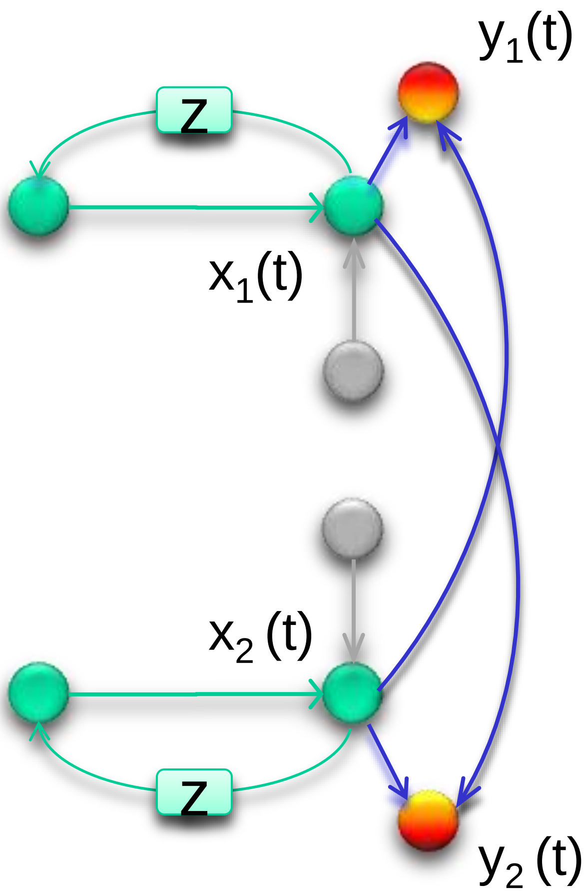
Structural / G - Causality

Mixed outputs: EEG

○ In some instances it is the physical process of observation that separates us from the time-series of interest. For example cortical sources and scalp based sensors (the mixing problem).



DGP Symbolic representation



Stochastic equivalence

○ It is also possible that there is both a non-diagonal observation matrix **and** covariate noise but these situations correspond to stochastically equivalent DGPs and cannot be disambiguated without further assumptions

$$y_i = \sum_{k=1}^{K} A_{P,k} y_{i-k} + b + R(\theta) S u_i$$

$$x_i = \sum_{k=1}^{K} R^T A_{P,k} R x_{i-k} + R^T b + S u_i$$

$$y = R^T x$$

R is a rotation matrix

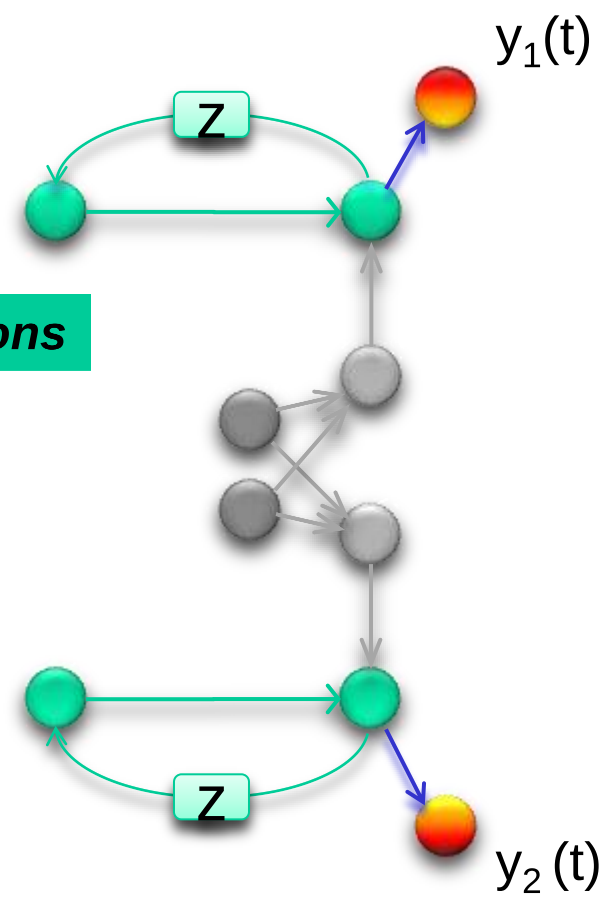
S is a diagonal (scaling) matrix

Covariate innovations

$\equiv_*$

Mixed output

DGP Symbolic representation



Stochastic equivalence

○ It is also possible that there is both a non-diagonal observation matrix **and** covariate noise but these situations correspond to stochastically equivalent DGPs and cannot be disambiguated without further assumptions

$$y_i = \sum_{k=1}^{K} A_{P,k} y_{i-k} + b + R(\theta) S u_i$$

$$x_i = \sum_{k=1}^{K} R^T A_{P,k} R x_{i-k} + R^T b + S u_i$$

$$y = R^T x$$

R is a rotation matrix

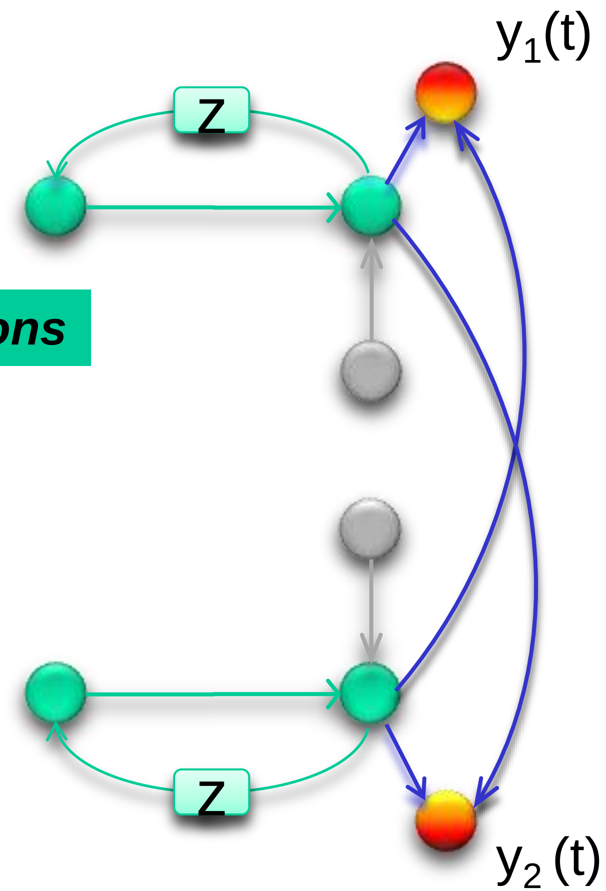
S is a diagonal (scaling) matrix

DGP Symbolic representation

Covariate innovations

$\equiv_*$

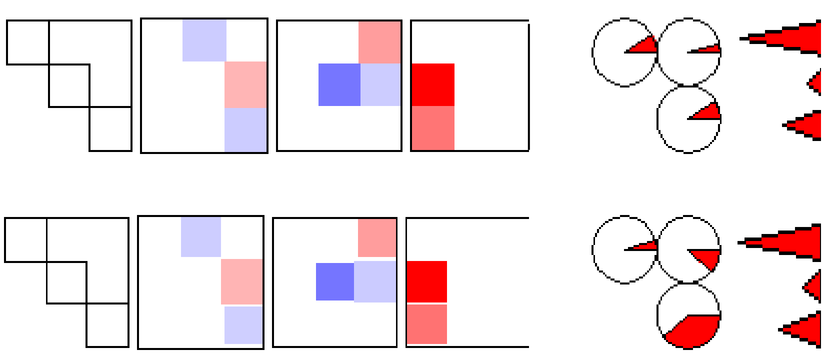
Mixed output



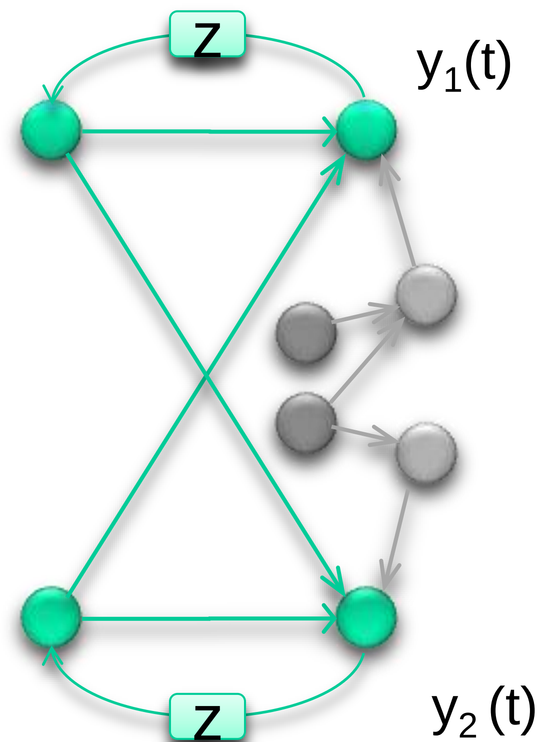
Noise covariance estimation

○ Instantaneous mixing / innovations covariance can be used to establish 'source' causality (Moneta 2008), (to follow!)

○ If a triangular structure is imposed on the instantaneous 'mixing' matrix of a linear SVAR the estimate of the equivalent noise covariance is unbiased (Popescu, 2008)



DGP Symbolic representation



Structural / G - Causality

Data Generating Process

$$x_n = A_{U,0} x_n + \sum_{i=1}^{K} A_{U,i} x_{n-i} + b_U + S_U e_{U,n}$$

strictly upper diagonal: $A_{U,0,p,q} = 0 \quad if \quad q \le p$

Zero-lag AR system

Can be solved by standard 2-norm linear **regression**
Strictly upper diagonal means resulting residuals **are not correlated.**

$$\boxed{S_U^{-1}(I - A_{U,0})} x_n = S_U^{-1} \sum_{i=1}^{K} A_{U,i} x_{n-i} + S_U^{-1} b_U + e_{U,n}$$

$$\boxed{U_U \Sigma_U V_U^T} x_n = S_U^{-1} \sum_{i=1}^{K} A_{U,i} x_{n-i} + S_U^{-1} b_U + e_{U,n}$$

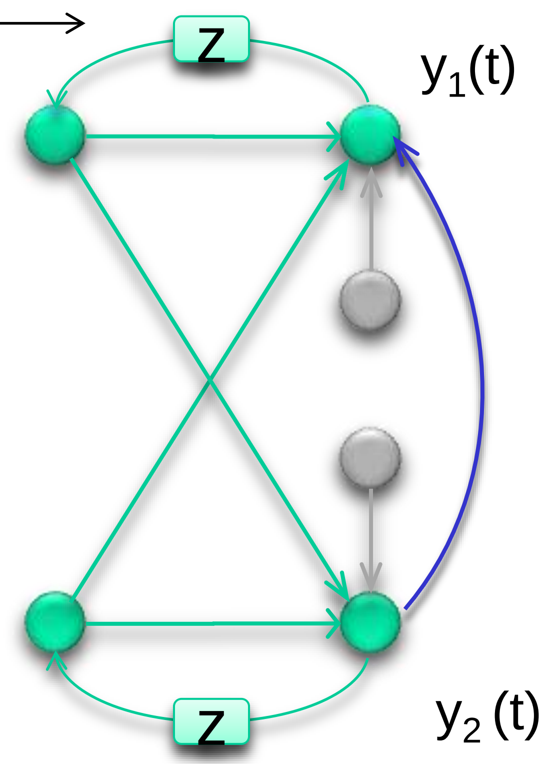
$$V_U^T x_U = \Sigma_U^{-1} U_U^T S_U^{-1} \sum_{i=1}^{K} A_{U,i} x_{n-i} + \Sigma_U^{-1} U_U^T S_U^{-1} b_U + \Sigma_U^{-1} \left(U_U^T e_{U,n} \right)$$

$$y_n = \boxed{V^T} x_n$$

$$y_n = \sum_{i=1}^{K} \boxed{\Sigma_U^{-1} U_U^T S_U^{-1} A_{U,i} V_U} y_{n-i} + \boxed{\Sigma_U^{-1} U_U^T S_U^{-1}} b_U + \boxed{\Sigma_U^{-1}} e_n$$

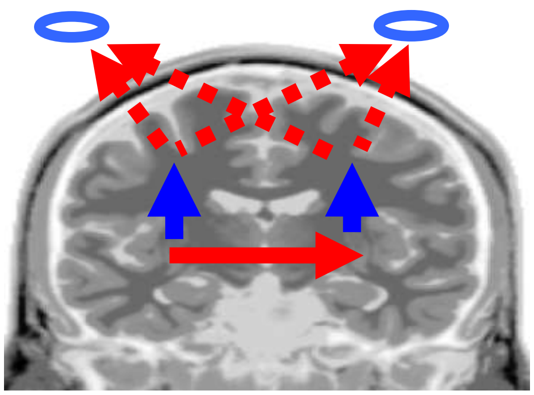
Mixed output

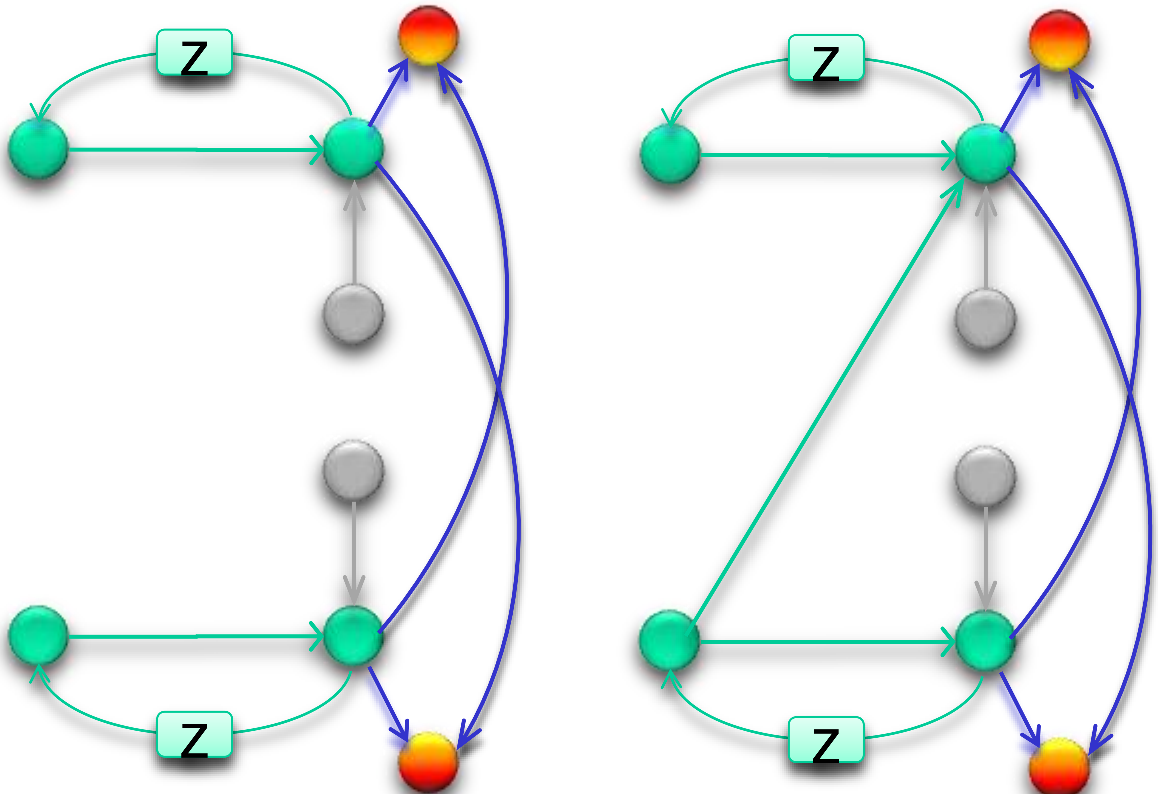
DGP Symbolic representation



Phase slope index

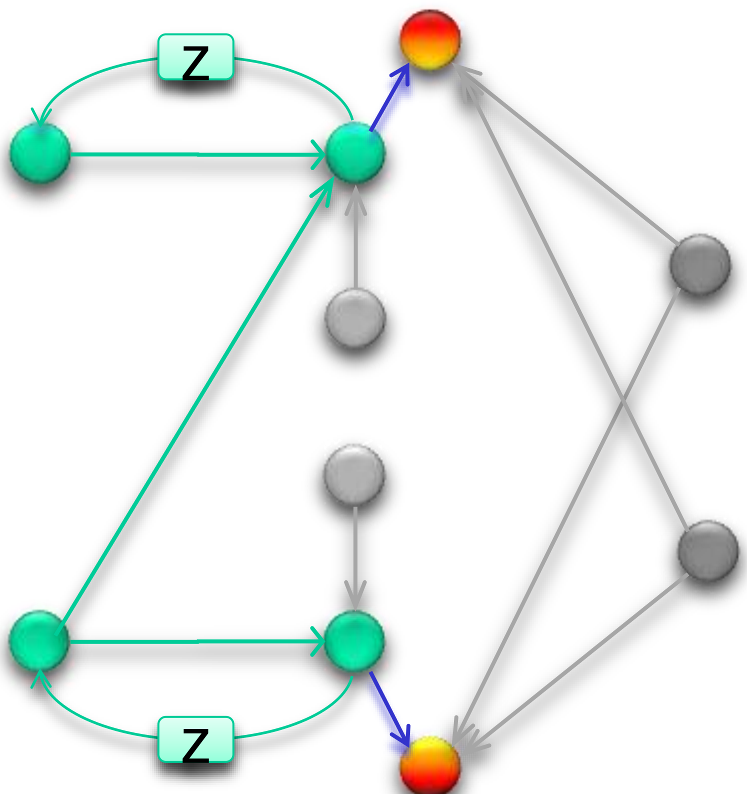
Basic principle: mixing does not affect the imaginary part of the complex coherency of a multivariate time series (Nolte 2004)





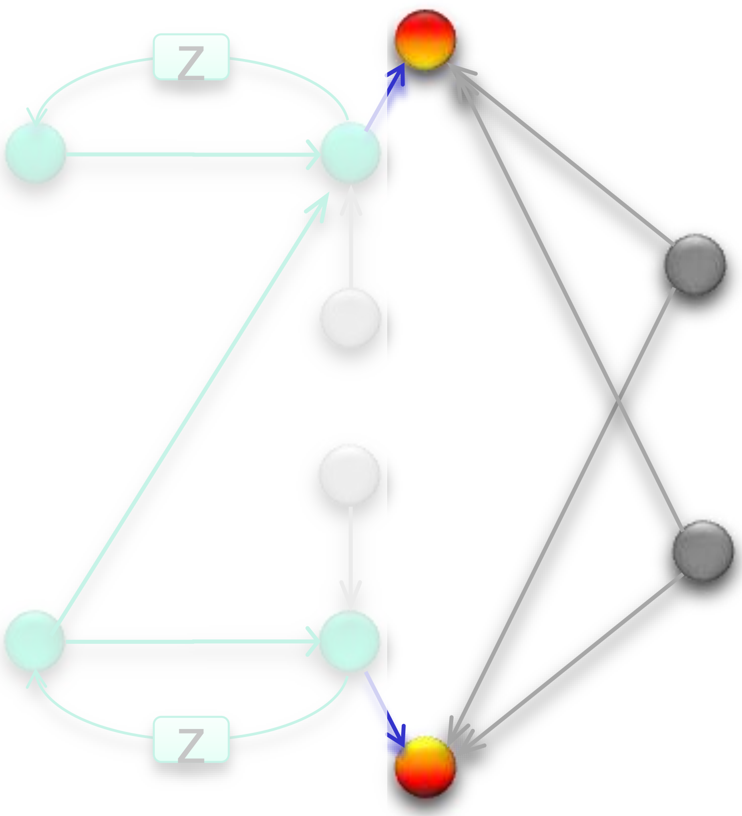
Phase slope index

Let us consider the case of a dynamically interacting system with correlated noise observations



Phase slope index

Let us consider the case of a dynamically interacting system with correlated noise observations. Relative influence of covariate noise?

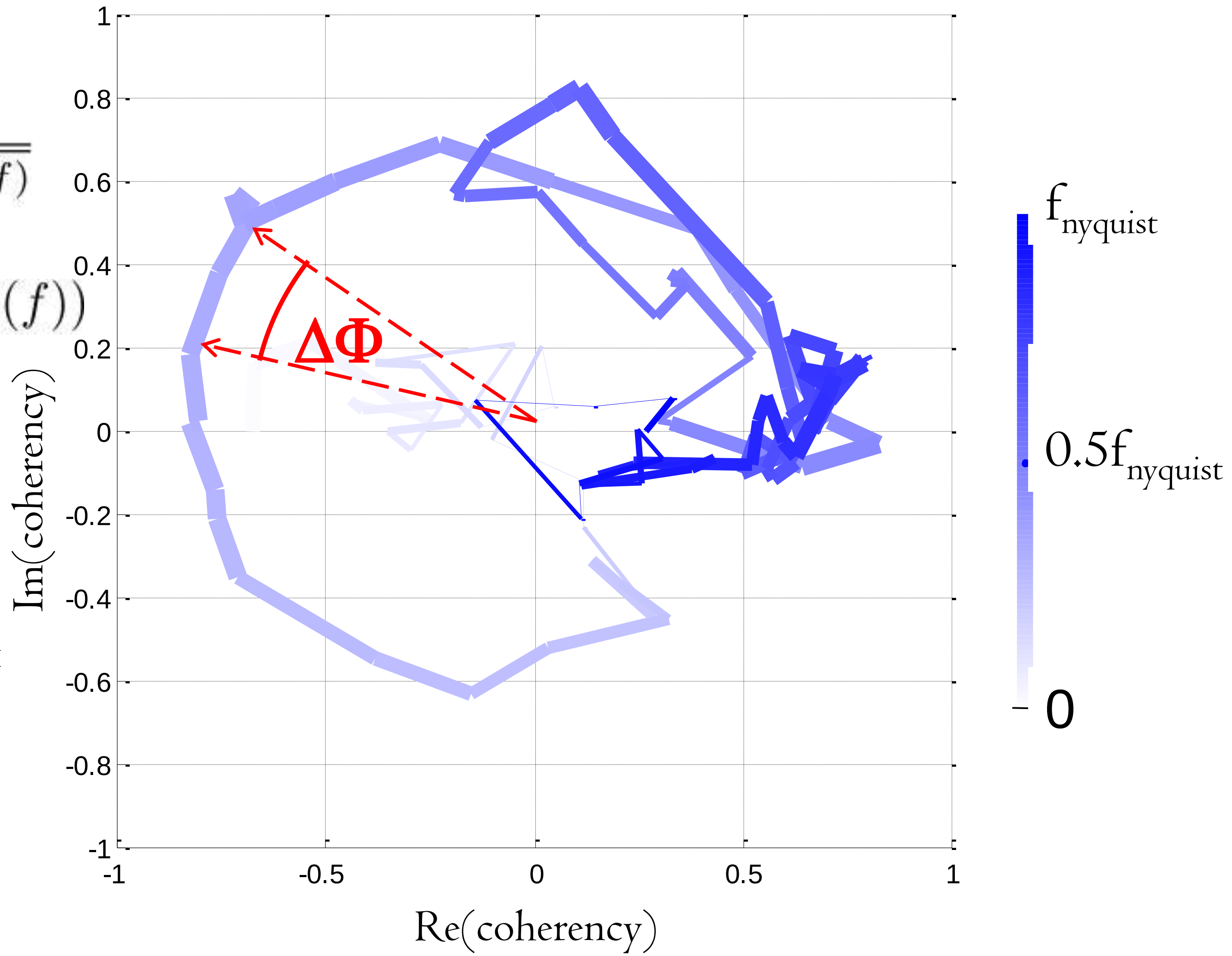


Phase Slope Index

$$C_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}$$

$$C_{ij}(f) = \alpha_{ij}(f)\exp(i\Phi(f))$$

Complex coherency C_{ij} is calculated from the complex spectral density S_{ij}

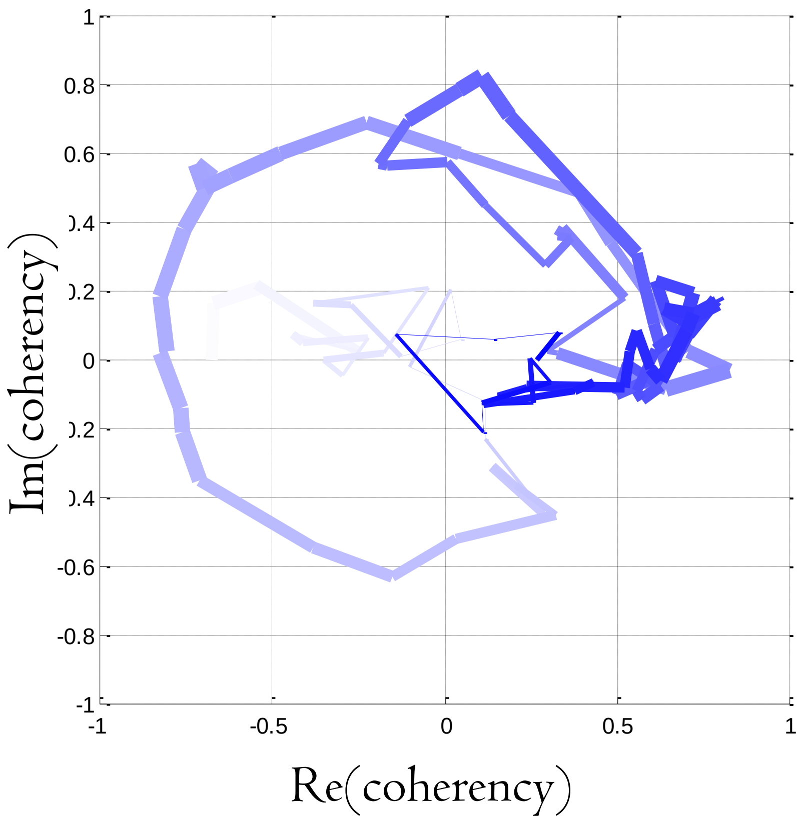


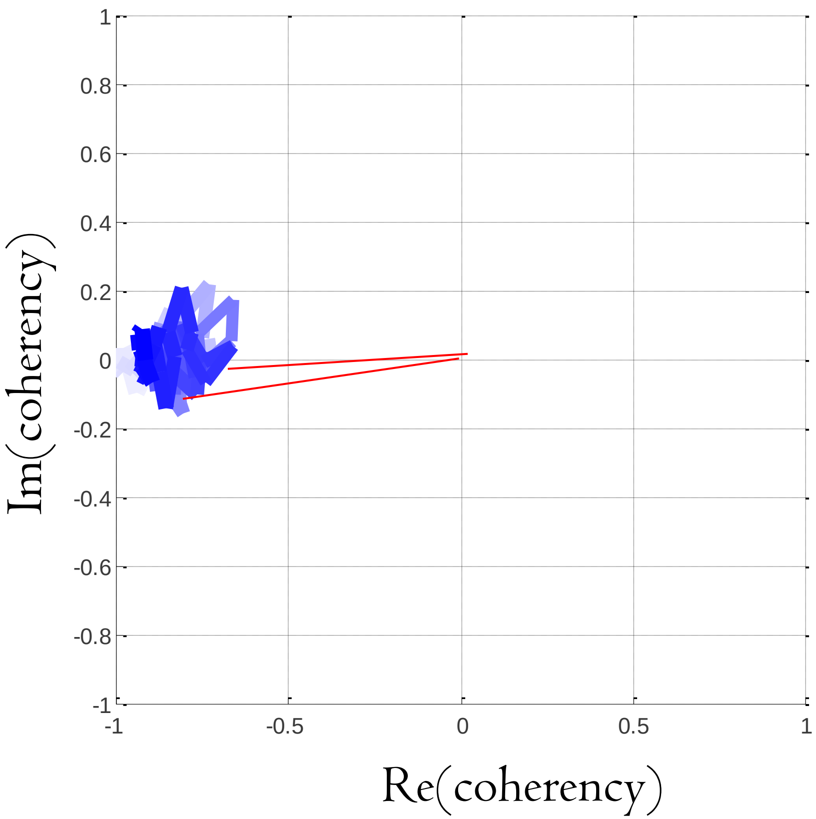
Phase slope index

$$\tilde{\Psi}_{ij} = \sum_{f \in F} \alpha_{ij}(f)\alpha_{ij}(f+\delta f)\sin(\Phi(f+\delta f) - \Phi(f))$$

$$\tilde{\Psi}_{ij} \approx \sum_{f \in F} \alpha_{ij}(f)\alpha_{ij}(f+\delta f)(\Phi(f+\delta f) - \Phi(f))$$

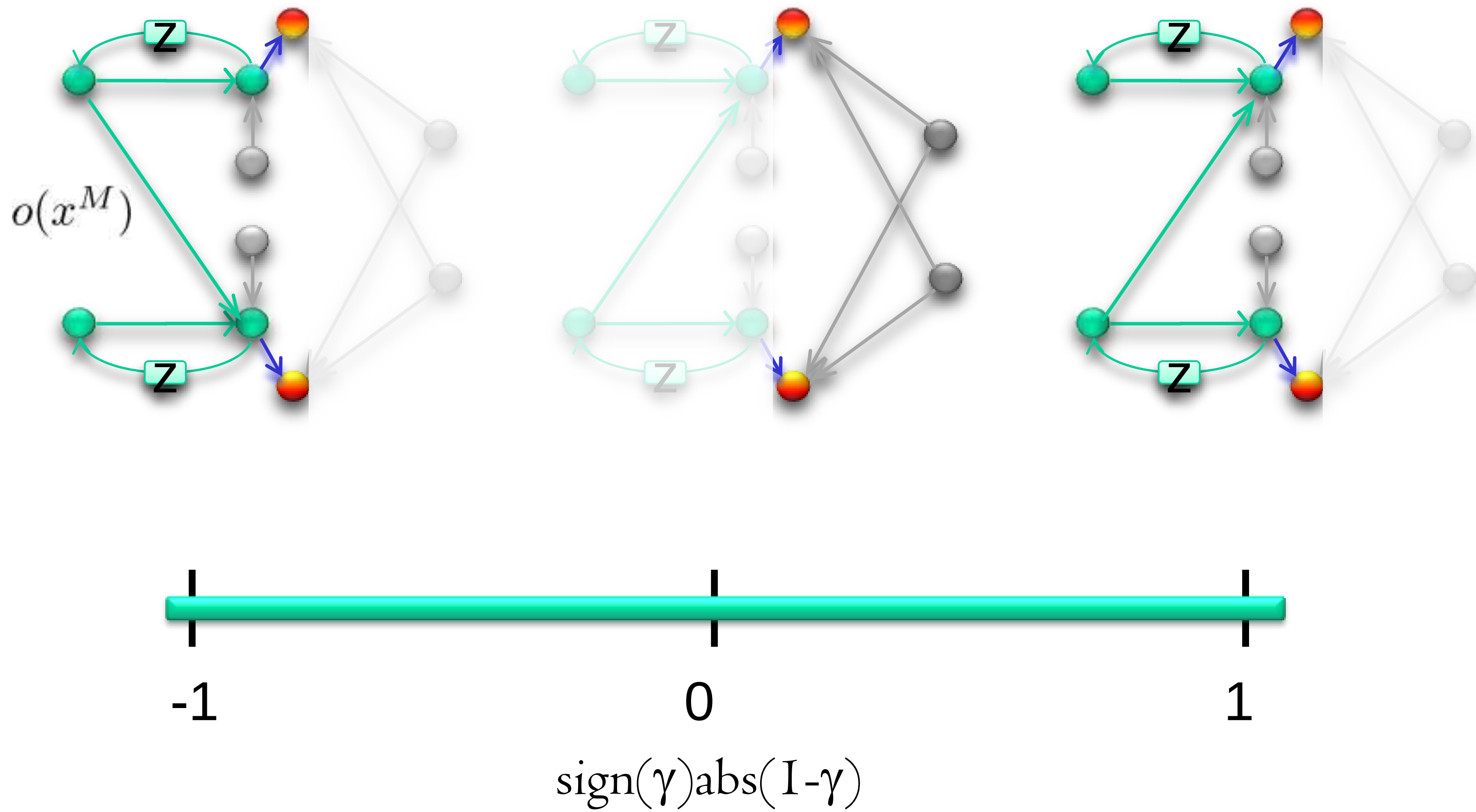
$$\tilde{\Psi}_{ij} \approx \sum_{f \in F} \alpha_{ij}(f)^2(\Phi(f+\delta f) - \Phi(f))$$



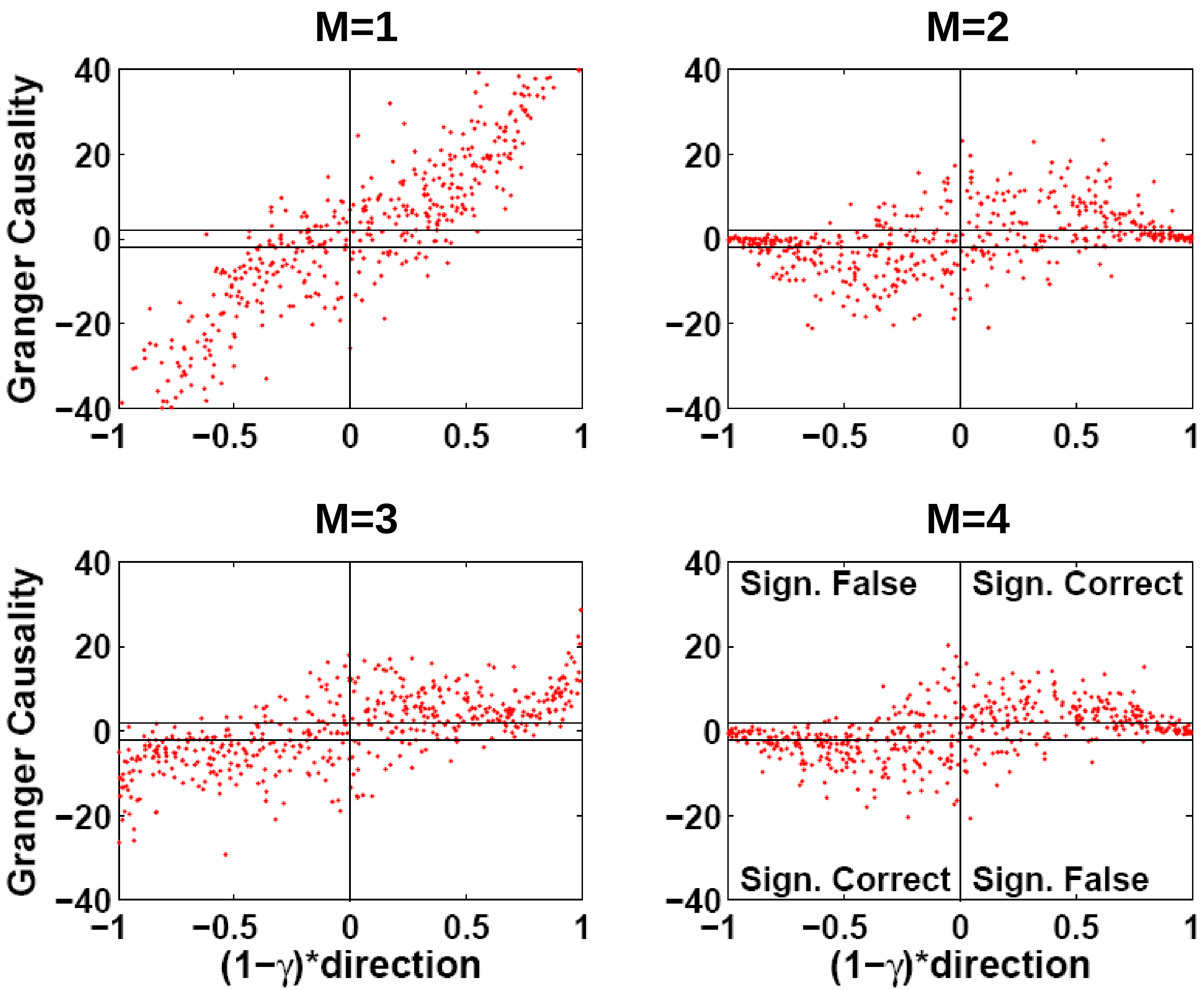


PSI: benchmark data

Causality workbench: 1000 simulations of linear/**nonlinear** systems

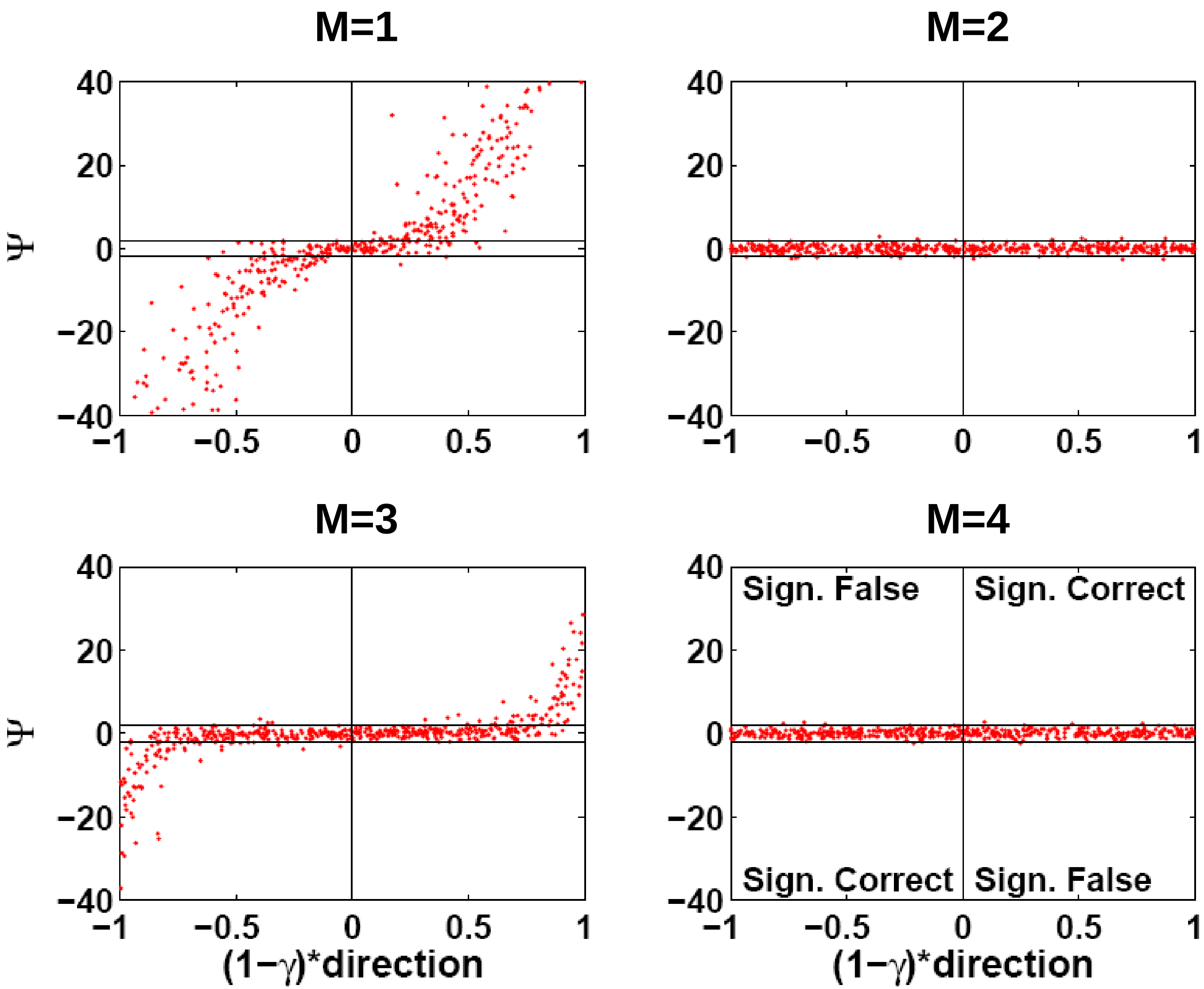


G-Causality Results



- 10th order AR models used
- **M** is the polynomial degree of the nonlinear coupling term

PSI Results


M=1
M=2
M=3
M=4
40
20
0
−20
−40
−1
−0.5
0
0.5
1
Ψ
Sign. False
Sign. Correct
Sign. Correct
Sign. False
(1−γ)*direction
(1−γ)*direction

Conclusions

- PSI does offer some advantages of Granger causality or model based method determination.
- PSI and related extensions can give statistical estimates of causality, dependence and non-causality and is conservative.
- Model based methods are limited by limitations in modeling technique: too few parameters may miss interactions, too many will over-fit, covariate innovations and AR coefficients are difficult to co-estimate.
- Future developments require DAG/ acyclic causal graph inference in multivariate time series.
- Complex non-stationarities not yet addressed.

Acknowledgments

PSI:

G. Nolte [1], A. Ziehe [1], N. Krämer [2], K.R.-Müller [2], Vadim V. Nikulin [3], Alois Schlögl [4], Tom Brismar [5]

1 Fraunhofer Institute FIRST, Berlin , [2] TU Berlin, [3] Charite Klinikum Berlin, [4] TU Graz, [5] Karolinska Institutet, Stockholm

PSI REFERENCES:

Nolte G, Ziehe A, Nikulin VV, Schlögl A, Krämer N, Brismar T, Müller KR. "Robustly estimating the flow direction of information in complex physical systems." Physical Review Letters 00(23):234101 . 2008.

Nolte G, Ziehe A, Krämer N, Popescu F, and Müller KR, "Comparison of Granger causality and phase slope index," Journal of Machine Learning Research Workshop and Conference Proceedings, in press., 2009.