Causalities and their Drivers in Synthetic and Financial Data

H. Ma[1,2], A. Haluszczynski[1], D. Prosperino[1], C. Räth[3]

[1] Allianz Global Investors, Seidlstraße 24-24a, 80335 München

[2] Ludwig-Maximilians-Universität, Fakultät für Physik, Schellingstraße 4, 80799 München

[3] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für KI Sicherheit, Sankt Augustin & Ulm

Abstract

Identifying and describing the dynamics of complex systems is a central challenge in various areas of science, such as physics, finance, or climatology. Here, we analyze the causal structure of chaotic systems using Fourier transform surrogates that enables us to identify the different (linear and nonlinear) causality drivers.

We further show that a simple rationale and calibration algorithm are sufficient to extract the governing equations directly from the causal structure of the data.

We demonstrate the applicability of the framework to real-world dynamical systems using financial data (stock indices from Europe, United States, China, Emerging Markets, Japan and Pacific excluding Japan) before and after the COVID-19 outbreak. It turns out that the pandemic triggered a fundamental rupture in the world economy, which is reflected in the causal structure and the resulting equations. Specifically, nonlinear causal relations have significantly increased in the global financial market after the COVID-19 outbreak [1].

Data

Synthetic data: Simulations of standard three-dimensional chaotic systems (Lorenz, Halvorsen) with tunable nonlinearities.

Financial data: Six major economies and their corresponding MSCI stock indices between November 2018 and May 2021: Europe (EU), United States (US), China (CN), Emerging Markets (EE), Japan (JP), and Pacific excluding Japan (PX).

Daily closing prices are converted to log returns.

Surrogates

In order to disentangle the contribution from linear and nonlinear correlations to the causal measures, we create surrogate time series which exhibit exactly the same linear properties as the original time series whereas nonlinear properties are fully randomized. This can be achieved by using *Fourier Transform (FT) surrogates* [2]

Causality Measures

We use the three commonly known causality measures:

- **Transfer Entropy (TE)** [3]
- **Convergent Cross Mapping (CCM)** [4]
- **Granger Causality (GC)** [5]

GC only serves as a verification for our analysis, since it is based on autoregression and should, therefore, only capture causality arising from linear properties.

Causalities are estimated between the time series and its corresponding surrogates:

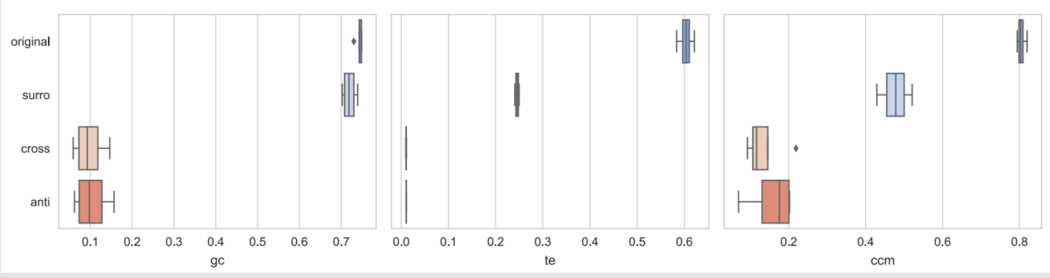
$$\psi^{surro}(\boldsymbol{x},\boldsymbol{y}) \equiv \frac{1}{K}\sum_{k=1}^{K}\psi\left(\tilde{\boldsymbol{x}}^{(k)},\tilde{\boldsymbol{y}}^{(k)}\right)$$

$$\psi^{cross}(\boldsymbol{x},\boldsymbol{y}) \equiv \frac{1}{K}\sum_{k=1}^{K}\psi\left(\tilde{\boldsymbol{x}}^{(k)},\boldsymbol{y}\right) \qquad \psi^{anti}(\boldsymbol{x},\boldsymbol{y}) \equiv \frac{1}{K}\sum_{k=1}^{K}\psi\left(\boldsymbol{x},\tilde{\boldsymbol{y}}^{(k)}\right)$$

References

1) H. Ma et al., Chaos, **32**, 103128 (2022)
2) C. Räth et al., PRL, **109**, 144101 (2012)
3) T. Schreiber, PRL, **85**, 461 (2000)
4) G. Sugihara et al., Science, **338**, 496 (2012)
5) C. Granger, *Essays in Econometrics: Collected Papers of Clive W.J. Granger* (Cambridge University Press, 2001), Vol. 32

Causalities for the Lorenz System



The surrogate analysis reveals that a significant amount of the causalities stems from the nonlinear correlations among the time series.

Deriving Governing Equations

By separating linear and nonlinear causalities, the terms of the governing equations become separately deducible, thus:

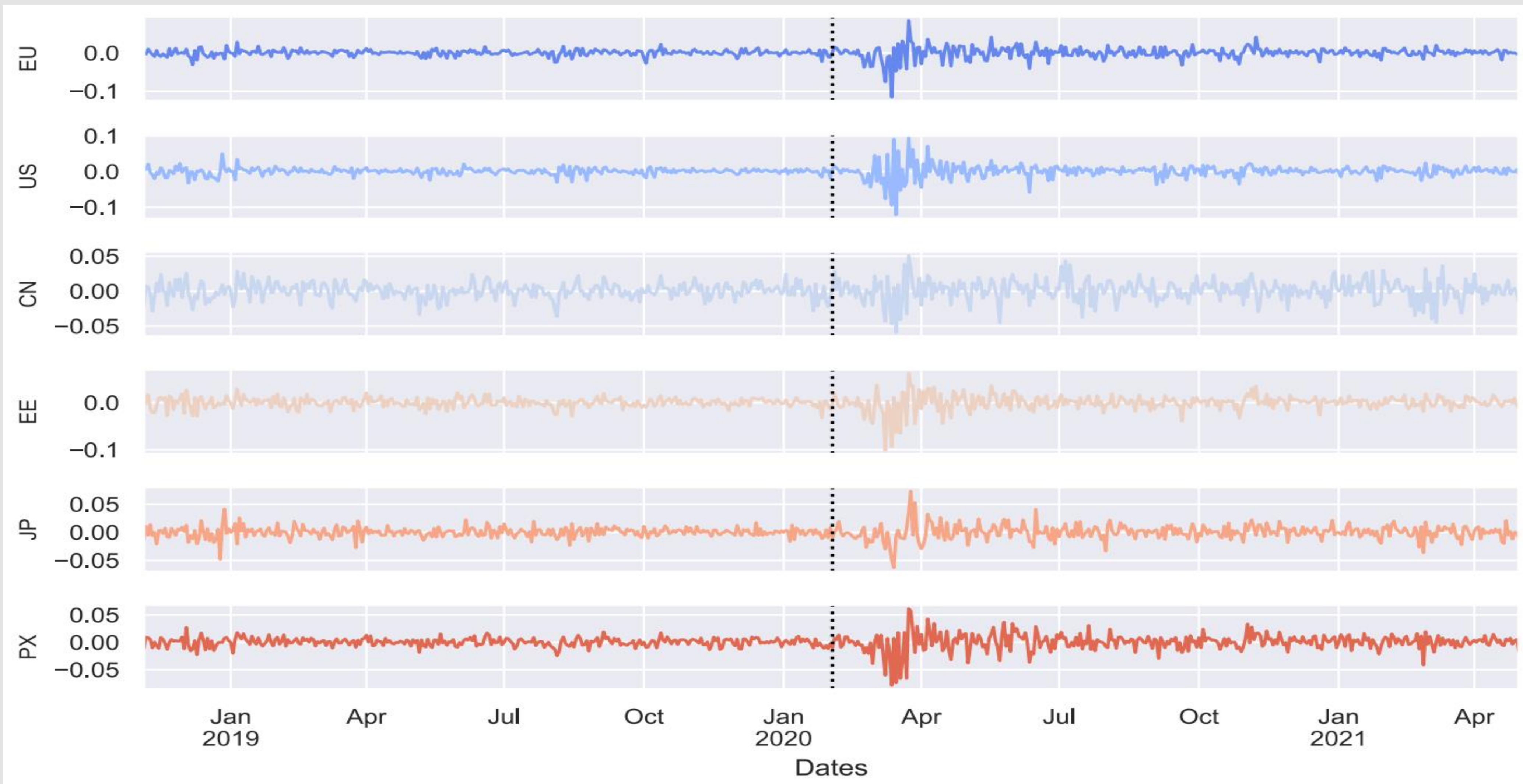
$$\frac{d\boldsymbol{x}}{dt} = \left(\frac{d\boldsymbol{x}}{dt}\right)_{lin} + \left(\frac{d\boldsymbol{x}}{dt}\right)_{nl} = \boldsymbol{\Psi}^{lin}\boldsymbol{x} + \boldsymbol{\Psi}^{nl}\odot\boldsymbol{x}^{n} \qquad \boldsymbol{\Psi}^{lin} = \delta_{i,j}\boldsymbol{\Psi}^{cross}_{i,j} + (1-\delta_{i,j})\,\boldsymbol{\Psi}^{surro}_{i,j}$$

with

$$\left(\frac{dx_j}{dt}\right)_{nl} = \Theta\left(\Psi^{nl}_{i,j} - 2\theta\right)x_i^2 \qquad \left(\frac{dx_j}{dt}\right)_{nl} = \sum_{i=k}^{n}\sum_{j\le i}^{l}\Theta\left(\Psi^{nl}_{i,j} + \Psi^{nl}_{j,i} - 4\theta\right)x_i x_j$$

For the standard chaotic systems the governing equations can be perfectly deduced.

Financial Data during Corona Outbreak:



Governing equations before and after Corona Outbreak

Economy	Before outbreak linear	Before outbreak nonlinear	After outbreak linear	After outbreak nonlinear
$\frac{dx_{eu}}{dt}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{eu}\,x_{px}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{px}$	$x_{jp}x_{px}+x_{us}x_{cn}+x_{us}x_{ee}+x_{us}x_{jp}+x_{cn}x_{ee}+x_{cn}x_{jp}+x_{ee}x_{jp}+x_{ee}x_{px}+x_{ee}x_{px}+x_{cn}x_{px}+x_{us}x_{px}$
$\frac{dx_{us}}{dt}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{cn}\,x_{px}$	$x_{eu}+x_{us}+x_{cn}+x_{jp}+x_{px}$	$x_{cn}x_{px}+x_{ee}x_{px}+x_{cn}x_{ee}$
$\frac{dx_{cn}}{dt}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{eu}\,x_{us}$	$x_{cn}+x_{ee}+x_{px}$	$x_{eu}x_{ee}+x_{us}x_{ee}+x_{eu}x_{us}+x_{eu}x_{px}+x_{ee}x_{px}+x_{us}x_{px}$
$\frac{dx_{ee}}{dt}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{jp}\,x_{px}$	$x_{eu}+x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{eu}x_{cn}+x_{us}x_{cn}+x_{jp}x_{px}+x_{us}x_{jp}+x_{cn}x_{jp}+x_{eu}x_{jp}+x_{eu}x_{us}+x_{eu}x_{px}+x_{cn}x_{px}+x_{us}x_{px}$
$\frac{dx_{jp}}{dt}$	$x_{cn}+x_{jp}+x_{px}+x_{eu}x_{ee}$	$x_{eu}x_{cn}+x_{us}x_{cn}+x_{us}x_{ee}+x_{cn}x_{jp}+x_{eu}x_{us}+x_{eu}x_{px}+x_{ee}x_{px}+x_{cn}x_{px}+x_{us}x_{jp}$	$x_{us}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{us}x_{ee}+x_{eu}x_{us}+x_{eu}x_{ee}$
$\frac{dx_{px}}{dt}$	$x_{eu}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{us}\,x_{px}$	$x_{eu}+x_{cn}+x_{ee}+x_{jp}+x_{px}$	$x_{us}x_{ee}+x_{eu}x_{us}+x_{eu}x_{ee}$

We find that all economies except Japan have only one nonlinear term before the February 2020 COVID-19 pandemic outbreak. In contrast, the equations for the post-pandemic outbreak phase have at least three nonlinear terms in all economies, suggesting that nonlinearity has increased in the global financial market.

Discussion

We analyzed the linear and nonlinear causal relations between variables in dynamical systems using different inference techniques and Fourier transform surrogates, which filter out the nonlinear properties of time series. We find for several chaotic systems that nonlinearity is a key driver of causality.

Furthermore, we developed a constructive and fully transparent rationale to derive the correct governing equations of chaotic systems directly from their causal structures - the resulting ease of interpretation of this approach is the main advantage in comparison to black-box machine learning methods.

Finally, we applied our methods to stock indices from different economies and found that the outbreak of the COVID- 19 pandemic triggered a structural change in the global financial markets.

DLR Deutsches Zentrum für Luft- und Raumfahrt