Statistica Sinica Preprint No: SS-2023-0317

Title	Statistical Inference for Local Granger Causality
Manuscript ID	SS-2023-0317
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0317
Complete List of Authors	Yan Liu, Masanobu Taniguchi and Hernando Ombao
Corresponding Authors	Yan Liu
E-mails	liu@waseda.jp
Notice: Accepted author version.	

Statistical Inference for Local Granger Causality

Yan Liu[1], Masanobu Taniguchi[1], Hernando Ombao[2]

[1] *Faculty of Science and Engineering, Waseda University, Japan*

[2] *Statistics Program, King Abdullah University of Science and Technology, Saudi Arabia*

Abstract:

Granger causality has been employed to investigate causality relations between components of stationary multiple time series. We generalize this concept by developing statistical inference for local Granger causality for multivariate locally stationary processes. Our proposed local Granger causality approach captures time-evolving causality relationships in nonstationary processes. The proposed local Granger causality is well represented in the frequency domain and estimated based on the parametric time-varying spectral density matrix using the local Whittle likelihood. Under regularity conditions, we demonstrate that the estimators converge to multivariate normal in distribution. Additionally, the test statistic for the local Granger causality is shown to be asymptotically distributed as a quadratic form of a multivariate normal distribution. For practical demonstration, the proposed local Granger causality method uncovered new functional connectivity relationships between channels in brain signals. Moreover, the method was applicable to topological data analysis to identify structural changes in financial data.

Key words and phrases: Brain signals, Local Granger causality, Local Whittle likelihood, Multivariate locally stationary processes, Time-varying spectral density matrix, Topological data analysis

1. Introduction

Statistical inference for cause and effect remains at the forefront of many studies including biology, medicine, physical systems, environmental science, public health, policy and finance. However, there remain challenges on inference because causality is notoriously difficult to establish. Granger causality, proposed in Granger (1963) and Granger (1969), is a milestone of causal inference in dynamic models. In broad terms, Granger causality from a time series $\{Y_t\}$ to another series $\{X_t\}$ measures the predictive ability from the series $\{Y_t\}$ to $\{X_t\}$. If the predictive ability of $(X_s, Y_s)_{s<t}$ on X_t is not different from the predictive ability of $(X_s)_{s<t}$ on X_t, then there is "no Granger causal relationship" from the series $\{Y_t\}$ to $\{X_t\}$. Thus, Granger causality analysis is important for determining whether or not a set of variables contains useful information for improving the prediction of another set of variables. Measures of linear dependence and feedback between components of a multivariate time series in both time and frequency domains have been considered in Geweke (1982) and Geweke (1984). In Hosoya (1991), a refinement of the above measures was proposed, which also has a well-defined representation in the frequency domain.

The inference for Granger causality using ARIMA modeling for nonstationary processes dates back to Sims et al. (1990). The paper elucidated nonstandard limiting distributions for Granger causality tests. Toda and Phillips (1993, 1994) considered the Wald test for cointegrated processes and derived that the test statistic is asymptotically distributed as a χ^2-distribution under the null hypothesis. Granger and Lin (1995) considered the Granger causality for nonstationary bivariate cointegrated processes. Other theoretical considerations can be found in Dolado and Lütkepohl (1996); Yamada and Toda (1998), to name a few. A numerical comparison for ARIMA models has been investigated in Clarke and Mirza (2006). A thorough treatment of the Granger causality for multivariate time series was discussed in Lütkepohl (2005).

In this paper, we propose a local Granger causality (LGC) measure based on the locally stationary process. A locally stationary process has a Cramér-like representation but its transfer function is allowed to change over time. Formal models for the time-varying spectra of nonstationary processes have been developed since this concept was introduced in Priestley (1965). A more theoretically rigorous framework for multivariate locally stationary processes have been formulated in Dahlhaus (2000). To estimate the time-varying spectral density of a locally stationary process, Neumann and von Sachs (1997) developed a wavelet estimator based on the pre-periodogram. The parameter estimation for an evolutionary spectral was discussed in Dahlhaus and Giraitis (1998). Local inference for locally stationary time series was investigated in Dahlhaus (2009). Moreover, Dahlhaus and Polonik (2009) constructed the estimation theory for the weak convergence of the empirical spectral processes. To the best of our knowledge, despite the recent progress on models that capture nonstationary behavior, local Granger causality has not yet been developed. To address this limitation, this paper undertakes the task of developing this local concept because many time series phenomena display Granger causality behavior that changes over the course of time (e.g., electroencephalograms and stock market indices). Thus, the contribution of this paper is a rigorous framework for statistical inference for LGC.

We focus on multivariate locally stationary processes to develop the statistical inference for LGC. Statistical inference for multivariate stationary processes has been discussed in Hannan (1970), Taniguchi and Kakizawa (2000), Shumway and Stoffer (2000) and references therein. A nonparametric method was developed in Taniguchi et al. (1996) to test the cross-relationships between multiple time series. The discriminant analysis for multivariate locally stationary processes based on the likelihood ratio was considered in Sakiyama and Taniguchi (2004). A SLEX model was proposed in Huang et al. (2004) to develop a discriminant scheme that can extract local features of time series. Several models were also considered for bivariate and multivariate nonstationary data using the SLEX basis which consists of well-localized Fourier-like waveforms (See Ombao et al. (2001) and Ombao et al. (2005)).

As noted, the goal of this paper is to develop statistical inference for LGC for multivariate locally stationary processes. In particular, LGC is expressed in the frequency domain using the foundational ideas on Granger causality for stationary processes. We develop a procedure for parameter estimation based on the local Whittle likelihood and derive the asymptotic distribution of the estimators. Under regularity conditions, the estimates are shown to converge to multivariate normal in distribution. The parameterized Granger causality, however, converges to normal or a quadratic form of normal random variables, which depends on the gradient of the causality measure. To illustrate the potential impact of the proposed LGC, we analyzed the log-returns of the financial data and multichannel electroencephalogram (EEG) data. Using the proposed method, the local Granger causality analyses produced insightful results such as new relationships between different brain signals and different topological structures between pre- and post-global financial crisis.

The remainder of the paper is organized as follows. In Section 2, we propose the local Granger causality. The properties of the local Granger causality are detailed immediately behind the definition. In Section 3, we develop statistical inference for local Granger causality based on the local Whittle estimation for multivariate locally stationary processes. In Section 4, we apply the proposed local Granger causality measure to EEG data and financial data. The proofs for the theoretical results are relegated to Section S3.

1.1 Notations

$O_{m\times M}$ denotes an $m \times M$ zero matrix; I_p denotes the $p \times p$ identity matrix; For any matrix A, let $\|A\|_\infty := \max_{1\le i\le p} \sum_{j=1}^{p} |a_{ij}|$. For a square matrix A, $|A|$ denotes its determinant. $\xrightarrow{d}$ denotes the convergence in distribution. Additionally, let l be a function such that

$$l(j) := \begin{cases} 1, & |j| \le 1, \\ |j| \ln^{1+\kappa}|j|, & |j| > 1, \end{cases} \tag{1.1}$$

for some constant $\kappa > 0$.

2. Local Granger Causality

In this section, we introduce the concept of local Granger causality in the framework of locally stationary processes. Let $\boldsymbol{X}_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(p)})^{\mathrm{T}}$ be a sequence of p-dimensional multivariate stochastic processes

$$\boldsymbol{X}_{t,T} = \sum_{j=-\infty}^{\infty} A_{t,T}(j)\boldsymbol{\epsilon}_{t-j}, \tag{2.2}$$

where the sequences $\{A_{t,T}(j)\}_{j\in\mathbb{Z}}$ satisfy the following conditions: there exists a positive constant C_A such that

$$\sup_{t,T}\|A_{t,T}(j)\|_\infty \le \frac{C_A}{l(j)}, \tag{2.3}$$

and there exists a sequence of continuous functions $A(\cdot, j) : [0,1] \to \mathbb{R}$ such that

(i) $\sup_u\|A(u,j)\|_\infty \le \frac{C_A}{l(j)}$;

(ii) $\sup_j\Big\|A_{t,T}(j) - A\Big(\frac{t}{T}, j\Big)\Big\|_\infty \le \frac{C_A}{l(j)}T^{-1}$;

(iii) $V\Big(\|A(\cdot,j)\|_\infty\Big) \le \frac{C_A}{l(j)}$,

where $V(f)$ is the total variation of the function f on the interval $[0,1]$, i.e., V is defined as

$$V(f) = \sup\Big\{\sum_{k=1}^{m}|f(x_k) - f(x_{k-1})|;\, 0 \le x_0 < \cdots < x_m \le 1,\, m \in \mathbb{N}\Big\}.$$

The process (2.2) is usually referred to as the multivariate locally stationary process. We impose the following assumptions on the process (2.2) for the estimation theory later on.

Assumption 1. *For the process in* (2.2)*, let* $\boldsymbol{\epsilon}_t$ *be independent and identically distributed with* $E\boldsymbol{\epsilon}_t = \mathbf{0}$ *and* $E\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^\top = \mathcal{K}$*, where the matrix* $\mathcal{K}$ *exists and all elements are bounded by* $C_{\mathcal{K}}$*. Furthermore, all elements in the* r*th moment of* $\boldsymbol{\epsilon}_t$ *exist and bounded by* $C_{\boldsymbol{\epsilon}}^{(r)}$*.*

Let m and M be two positive integers such that $p = m + M$. Suppose $\boldsymbol{X}_{t,T} = \big(\boldsymbol{X}_{t,T}^{(1)\top}, \boldsymbol{X}_{t,T}^{(2)\top}\big)^\top$, $\boldsymbol{X}_{t,T}^{(1)} \in \mathbb{R}^m$, $\boldsymbol{X}_{t,T}^{(2)} \in \mathbb{R}^M$, has the time-varying spectral density matrix $\boldsymbol{f}(u,\lambda)$ with the partition

$$\boldsymbol{f}(u,\lambda) = \begin{pmatrix} \boldsymbol{f}(u,\lambda)_{11} & \boldsymbol{f}(u,\lambda)_{12} \\ \boldsymbol{f}(u,\lambda)_{21} & \boldsymbol{f}(u,\lambda)_{22} \end{pmatrix} := \frac{1}{2\pi} A(u,\lambda)\mathcal{K}A(u,-\lambda)^\top, \qquad u \in [0,1], \tag{2.4}$$

where $A(u,\lambda) := \sum_{j=-\infty}^{\infty} A(u,j)\exp(\mathrm{i}j\lambda)$. Let $\Sigma(u)$ be the one-step-ahead prediction error covariance matrix based on the time-varying spectral density matrix $\boldsymbol{f}(u,\lambda)$ with the same partition. By the Kolmogorov's formula for multiple time series, we have

$$\det\Sigma(u) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\Big(\det 2\pi\boldsymbol{f}(u,\lambda)\Big)\,\mathrm{d}\lambda\right), \qquad \text{for any } u \in [0,1],$$

(see Hannan (1970), p.162).

Let $H(\tau_1,\tau_2) = \overline{\mathrm{sp}}(\boldsymbol{X}_{t,T}^{(1)}, 1 \le t \le \tau_1; \boldsymbol{X}_{t,T}^{(2)}, 1 \le t \le \tau_2)$ be the closed linear subspace generated by $\{\boldsymbol{X}_{t,T}^{(1)}, 1 \le t \le \tau_1; \boldsymbol{X}_{t,T}^{(2)}, 1 \le t \le \tau_2\}$. Especially, we use $H(\tau_1, 0)$ and $H(0,\tau_2)$ to express the closed linear subspace generated by $\{\boldsymbol{X}_{t,T}^{(1)}, t \le \tau_1\}$, $\{\boldsymbol{X}_{t,T}^{(2)}, t \le \tau_2\}$, respectively.

Introducing a companion process

$$\boldsymbol{Y}_{t,T}^{(2)} = \boldsymbol{X}_{t,T}^{(2)} - E\big(\boldsymbol{X}_{t,T}^{(2)} \mid H(t,t-1)\big), \tag{2.5}$$

we propose the local Granger causality measure from $\{\boldsymbol{X}_{t,T}^{(2)}\}$ to $\{\boldsymbol{X}_{t,T}^{(1)}\}$ as

$$\mathrm{GC}^{(2\to1)}(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{FGC}(u,\lambda)d\lambda, \tag{2.6}$$

where

$$\mathrm{FGC}(u,\lambda) = \ln \frac{|\boldsymbol{f}(u,\lambda)_{11}|}{\left|\boldsymbol{f}(u,\lambda)_{11} - 2\pi \boldsymbol{g}(u,\lambda)_{12}\tilde{\Sigma}(u)_{22}^{-1}\boldsymbol{g}(u,\lambda)_{21}\right|}.$$

Here, $\boldsymbol{g}(u,\lambda)$ is the time-varying spectral density matrix of the process $\{\big(\boldsymbol{X}_{t,T}^{(1)\top}, \boldsymbol{Y}_{t,T}^{(2)\top}\big)^\top\}$ (see Propositions 1 and 2 below), and $\tilde{\Sigma}(u)$ is an $(M\times M)$-matrix

$$\tilde{\Sigma}(u)_{22} = \Sigma(u)_{22} - \Sigma(u)_{21}\Sigma(u)_{11}^{-1}\Sigma(u)_{12}.$$

Remark 1. Let us briefly provide an interpretation of the local Granger causality measure $\mathrm{GC}^{(2\to1)}(u)$ in (2.6). Given a localized model around $u \in (0,1)$, i.e.,

$$\boldsymbol{X}_t(u) = \begin{pmatrix} \boldsymbol{X}_t^{(1)}(u) \\ \boldsymbol{X}_t^{(2)}(u) \end{pmatrix} = \sum_{j=-\infty}^{\infty} A(u,j)\boldsymbol{\epsilon}_{t-j},$$

the original Granger causality measure for stationary processes from $\{\boldsymbol{X}_t^{(2)}(u)\}$ to $\{\boldsymbol{X}_t^{(1)}(u)\}$ is defined as the log-ratio of the one-step-ahead prediction error based only on the process $\{\boldsymbol{X}_t^{(1)}(u)\}$ to that based on the process $\{\boldsymbol{X}_t(u)\}$ itself.

Mathematically speaking, the log-ratio in the original Granger causality measure is given by

$$\tilde{\mathrm{GC}}^{(2\to1)}(u) = \ln \frac{\det\,\mathrm{Cov}\Big(\boldsymbol{X}_{t+1}^{(1)}(u) - E\big(\boldsymbol{X}_{t+1}^{(1)}(u) \mid \boldsymbol{X}_t^{(1)}(u)\big)\Big)}{\det\,\mathrm{Cov}\Big(\boldsymbol{X}_{t+1}^{(1)}(u) - E\big(\boldsymbol{X}_{t+1}^{(1)}(u) \mid \boldsymbol{X}_t(u)\big)\Big)}.$$

The proposal of (2.6) follows from the Kolmogorov's formula, and incorporated the natural decomposition of frequency measure considered in Hosoya (1991).

As it has been seen, the proposal of the local Granger causality (2.6) is motivated by Hosoya's measure of causality (a term first coined in Granger and Lin (1995)) in combination with the nonstationary version of Kolmogorov's formula by Dahlhaus (1996). The companion process $\{\boldsymbol{Y}_{t,T}^{(2)}\}$ in (2.5) is introduced in order to remove the possible effect brought by the nonorthogonality between residuals of predictions $E\big(\boldsymbol{X}_{t,T}^{(1)} \mid H(t-1,t-1)\big)$ and $E\big(\boldsymbol{X}_{t,T}^{(2)} \mid H(t-1,t-1)\big)$.

We now start to explain properties of the proposed local Granger causality. Denote by $\Sigma_{t,T}$ the one-step-ahead prediction error covariance matrix, i.e.,

$$\Sigma_{t,T} = \operatorname{var}\big[\boldsymbol{X}_{t,T} - E\big(\boldsymbol{X}_{t,T} \mid H(t-1,t-1)\big)\big].$$

Proposition 1. *The companion process* $\{\boldsymbol{Y}_{t,T}^{(2)}\}$ *is a locally stationary process with the time-varying spectral density*

$$\boldsymbol{g}(u,\lambda)_{22} = \frac{1}{2\pi}\tilde{\Sigma}(u)_{22}. \tag{2.7}$$

Proof. From the definition of $\{\boldsymbol{X}_{t,T}\}$ in (2.2), we have

$$\boldsymbol{X}_{t,T} - E\big(\boldsymbol{X}_{t,T} \mid H(t-1,t-1)\big) = A_{t,T}(0)\boldsymbol{\epsilon}_t.$$

By the following formula (see Lemma 2.2 in Hosoya (1991)),

$$\begin{aligned}
&\boldsymbol{X}_{t,T}^{(2)} - E\big(\boldsymbol{X}_{t,T}^{(2)} \mid H(t,t-1)\big)\\
=&\big\{\boldsymbol{X}_{t,T}^{(2)} - E\big(\boldsymbol{X}_{t,T}^{(2)} \mid H(t-1,t-1)\big)\big\}\\
&-\Sigma_{t,T,21}\Sigma_{t,T,11}^{-1}\big\{\boldsymbol{X}_{t,T}^{(1)} - E\big(\boldsymbol{X}_{t,T}^{(1)} \mid H(t-1,t-1)\big)\big\},
\end{aligned}$$

and we find that

$$\boldsymbol{Y}_{t,T}^{(2)} = \boldsymbol{\epsilon}_t^{(2)} - \Sigma_{t,T,21}\Sigma_{t,T,11}^{-1}\boldsymbol{\epsilon}_t^{(1)} = \begin{pmatrix} -\Sigma_{t,T,21}\Sigma_{t,T,11}^{-1} & \boldsymbol{I}_M \end{pmatrix}\boldsymbol{\epsilon}_t.$$

In view of Example 2.3 (i) in Dahlhaus (2000), $\{\boldsymbol{Y}_{t,T}^{(2)}\}$ is locally stationary. A straightforward calculation gives the expression of $\boldsymbol{g}(u,\lambda)_{22}$ in (2.7). □

Let $\mathcal{H}^{(2)}(\tau)$ be the closed linear subspace generated by $\{\boldsymbol{Y}_{t,T}^{(2)}, 1 \le t \le \tau\}$. The Hosoya measure is defined as

$$\mathrm{HM}_{t,T}^{(2\to1)} := \ln \frac{\det \operatorname{var}\big[\boldsymbol{X}_{t,T}^{(1)} - E\big(\boldsymbol{X}_{t,T}^{(1)} \mid H(t-1,0)\big)\big]}{\det \operatorname{var}\big[\boldsymbol{X}_{t,T}^{(1)} - E\big(\boldsymbol{X}_{t,T}^{(1)} \mid \sigma\{H(t-1,0)\cup\mathcal{H}^{(2)}(t-1)\}\big)\big]}. \tag{2.8}$$

For any fixed $u \in [0,1]$, the time-varying spectral matrix $\boldsymbol{f}(u,\lambda)$ has a factorization

$$\boldsymbol{f}(u,\lambda) = \frac{1}{2\pi}\boldsymbol{\Lambda}(u,\mathrm{e}^{-\mathrm{i}\lambda})\boldsymbol{\Lambda}(u,\mathrm{e}^{\mathrm{i}\lambda})^*, \quad z \in \mathcal{D}, \tag{2.9}$$

(see Rozanov (1967)).

Proposition 2. *Suppose all eigenvalues of $A(u,\lambda)A(u,-\lambda)^\top$ are bounded from below by some constant $C > 0$ uniformly in u and λ, and all components of $A(u,\lambda)$ are differentiable in u and λ with bounded derivatives $(\partial/\partial u)(\partial/\partial\lambda)A(u,\lambda)_{ab}$ for $a,b \in \{1,2,\dots,p\}$. It holds that*

$$|\mathrm{GC}^{(2\to1)}(t/T) - \mathrm{HM}_{t,T}^{(2\to1)}| = o_t(1) + O_T(1),$$

where the $o_t(1)$ term is uniform in T and the $o_T(1)$ term is uniform in t.

Proof. From Proposition 1, we see that the process $\{(\boldsymbol{X}_{t,T}^{(1)\top}, \boldsymbol{Y}_{t,T}^{(2)\top})^\top\}$ is locally stationary. In view of Lemma 2.3 in Hosoya (1991), we see that the process has the time-varying spectral density matrix $\boldsymbol{g}(u,\lambda)$ with $\boldsymbol{g}(u,\lambda)_{11} = \boldsymbol{f}(u,\lambda)_{11}$ and

$$\begin{aligned} \boldsymbol{g}(u,\lambda)_{21} &= \boldsymbol{g}(u,-\lambda)_{12}^\top \\ &= \begin{pmatrix} -\Sigma(u)_{21}\Sigma(u)_{11}^{-1} & I_M \end{pmatrix} \boldsymbol{\Lambda}(u,0)\boldsymbol{\Lambda}(u,\mathrm{e}^{\mathrm{i}\lambda})^{-1} \begin{pmatrix} \boldsymbol{f}(u,\lambda)_{11} \\ \boldsymbol{f}(u,\lambda)_{12} \end{pmatrix}. \end{aligned} \tag{2.10}$$

A direct computation shows that the process $\{\boldsymbol{X}_{t,T}^{(1)} - E(\boldsymbol{X}_{t,T}^{(1)} \mid \sigma\{H(t-1,0) \cup \mathcal{H}^{(2)}(t-1)\})\}$ is still locally stationary and has the time-varying spectral density

$$\boldsymbol{g}(u,\lambda)_{11} - \boldsymbol{g}(u,\lambda)_{12}\boldsymbol{g}(u,\lambda)_{22}^{-1}\boldsymbol{g}(u,\lambda)_{21}.$$

Inspection of Theorem 3.2 in Dahlhaus (1996) for the nonstationary version of Kolmogorov's formula reveals that

$$\begin{aligned} &\det \operatorname{var}\big[\boldsymbol{X}_{t,T}^{(1)} - E\big(\boldsymbol{X}_{t,T}^{(1)} \mid H(t-1,0)\big)\big] \\ &\qquad = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\Big(\det 2\pi \boldsymbol{f}(t/T,\lambda)_{11}\Big)\,\mathrm{d}\lambda\right) + o_t(1) + o_T(1), \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} &\det \operatorname{var}\big[\boldsymbol{X}_{t,T}^{(1)} - E\big(\boldsymbol{X}_{t,T}^{(1)} \mid \sigma\{H(t-1,0) \cup \mathcal{H}^{(2)}(t-1)\}\big)\big] \\ &= \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\Big(\det 2\pi\big(\boldsymbol{g}(t/T,\lambda)_{11} - \boldsymbol{g}(t/T,\lambda)_{12}\boldsymbol{g}(t/T,\lambda)_{22}^{-1}\boldsymbol{g}(t/T,\lambda)_{21}\big)\Big)\,\mathrm{d}\lambda\right) \\ &\qquad + o_t(1) + o_T(1). \end{aligned} \tag{2.12}$$

Combining (2.11) and (2.12) yields the desired result. $\square$

The local Granger causality measure can be regarded as the limit of that constructed by the Wigner-Ville spectrum. To be specific, let $\boldsymbol{f}_{t,T}(\lambda)$ be the Wigner-Ville spectrum of the process $\{\boldsymbol{X}_{t,T}\}$, i.e.,

$$\boldsymbol{f}_{t,T}(\lambda) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \mathrm{Cov}\big(\boldsymbol{X}_{[t-s/2],T}, \boldsymbol{X}_{[t+s/2],T}\big)\exp(-\mathrm{i}\lambda s),$$

(see Martin and Flandrin (1985)). The measure of the Wigner-Ville spectrum now is

$$\mathrm{GC}_{t,T}^{(2\to 1)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{FGC}_{t,T}(\lambda)\,\mathrm{d}\lambda,$$

where

$$\mathrm{FGC}_{t,T}(\lambda) = \ln\frac{|\boldsymbol{f}_{t,T}(\lambda)_{11}|}{\left|\boldsymbol{f}_{t,T}(\lambda)_{11} - 2\pi\boldsymbol{g}_{t,T}(\lambda)_{12}\tilde{\Sigma}_{t,T,22}^{-1}\boldsymbol{g}_{t,T}(\lambda)_{21}\right|}$$

and $\tilde{\Sigma}(u)$ is an $(M\times M)$-matrix

$$\tilde{\Sigma}_{t,T,22} = \Sigma_{t,T,22} - \Sigma_{t,T,21}\Sigma_{t,T,11}^{-1}\Sigma_{t,T,12}.$$

Proposition 3. *Suppose $\boldsymbol{f}(u,\lambda)$ is uniformly Lipschiz continuous with respect to u and λ. For any sequence $t/T\to u$, We have*

$$\left|\mathrm{GC}^{(2\to 1)}(u) - \mathrm{GC}_{t,T}^{(2\to 1)}\right| = o(1).$$

Proof. We only show that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\big(\ln|\boldsymbol{f}(u,\lambda)_{11}| - \ln|\boldsymbol{f}_{t,T}(\lambda)_{11}|\big)\,\mathrm{d}\lambda = o(1),$$

to see the difference in the numerator. The denominator can be proved similarly.

Let us consider the scalar process $\boldsymbol{\alpha}^* \boldsymbol{X}_{t,T}^{(1)}$ with the Wigner-Ville spectrum $f_{t,T}^{\boldsymbol{\alpha}}(\lambda) := \boldsymbol{\alpha}^* \boldsymbol{f}_{t,T}(\lambda)_{11} \boldsymbol{\alpha}$ for any $\boldsymbol{\alpha} \in \mathbb{C}^m$. Comparing it with $f^{\boldsymbol{\alpha}}(u,\lambda) = \boldsymbol{\alpha}^* \boldsymbol{f}(u,\lambda)_{11} \boldsymbol{\alpha}$, by Theorem 2.2 in Dahlhaus (1996), we see that

$$\int_{-\pi}^{\pi} |f_{t,T}^{\boldsymbol{\alpha}}(\lambda) - f^{\boldsymbol{\alpha}}(u,\lambda)|^2 \,\mathrm{d}\lambda = o(1),$$

which implies that

$$\int_{-\pi}^{\pi} |f_{t,T}^{\boldsymbol{\alpha}}(\lambda) - f^{\boldsymbol{\alpha}}(u,\lambda)| \,\mathrm{d}\lambda = o(1), \tag{2.13}$$

since by the Cauchy-Schwarz inequality, we have

$$\int_{-\pi}^{\pi} |f_{t,T}^{\boldsymbol{\alpha}}(\lambda) - f^{\boldsymbol{\alpha}}(u,\lambda)| \,\mathrm{d}\lambda \le \sqrt{2\pi} \Big(\int_{-\pi}^{\pi} |f_{t,T}^{\boldsymbol{\alpha}}(\lambda) - f^{\boldsymbol{\alpha}}(u,\lambda)|^2 \,\mathrm{d}\lambda \Big)^{1/2}.$$

By Taylor's expansion, we have

$$\begin{aligned} \ln|\boldsymbol{f}_{t,T}(\lambda)_{11}| = \ln|\boldsymbol{f}(u,\lambda)_{11}| &+ \mathrm{Tr}\Big[\boldsymbol{f}(u,\lambda)_{11}^{-1}\big(\boldsymbol{f}_{t,T}(\lambda)_{11} - \boldsymbol{f}(u,\lambda)_{11}\big)\Big] \\ &+ o\Big(\mathrm{Tr}\Big[\boldsymbol{f}(u,\lambda)_{11}^{-1}\big(\boldsymbol{f}_{t,T}(\lambda)_{11} - \boldsymbol{f}(u,\lambda)_{11}\big)\Big]\Big). \end{aligned} \tag{2.14}$$

Remembering that $\boldsymbol{f}(u,\lambda) = \frac{1}{2\pi}\boldsymbol{\Lambda}(u, \mathrm{e}^{-\mathrm{i}\lambda})\boldsymbol{\Lambda}(u,\mathrm{e}^{\mathrm{i}\lambda})^*$ from (2.9), we see that there exists an $m \times m$ Hermitian matrix $\boldsymbol{B}$ such that $\boldsymbol{f}(u,\lambda)_{11}^{-1} = \boldsymbol{B}^*\boldsymbol{B}$, and thus

$$\mathrm{Tr}\Big[\boldsymbol{f}(u,\lambda)_{11}^{-1}\big(\boldsymbol{f}_{t,T}(\lambda)_{11} - \boldsymbol{f}(u,\lambda)_{11}\big)\Big] = \mathrm{Tr}\Big[\boldsymbol{B}\big(\boldsymbol{f}_{t,T}(\lambda)_{11} - \boldsymbol{f}(u,\lambda)_{11}\big)\boldsymbol{B}^*\Big], \tag{2.15}$$

which is a sum of quadratic forms $f_{t,T}^{\boldsymbol{\alpha}}(\lambda) - f^{\boldsymbol{\alpha}}(u,\lambda)$. Applying (2.13) to (2.15) yields

$$\int_{-\pi}^{\pi} \mathrm{Tr}\Big[\boldsymbol{f}(u,\lambda)_{11}^{-1}\big(\boldsymbol{f}_{t,T}(\lambda)_{11} - \boldsymbol{f}(u,\lambda)_{11}\big)\Big] \,\mathrm{d}\lambda = o(1),$$

and by observing (2.14), we conclude that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|\boldsymbol{f}(u,\lambda)_{11}| - \ln|\boldsymbol{f}_{t,T}(\lambda)_{11}| \,\mathrm{d}\lambda = o(1).$$

□

Remark 2. The construction of a linear predictor in practice for the locally stationary process may be of interest to some readers. It can be shown that the predictor for the locally stationary process and that for the stationary approximation are asymptotically equivalent under adequate conditions. In contrast, we focus on the nonstationary version of Kolmogorov's formula found in Dahlhaus (1996). We elucidated that our local causality measure, as the limit of the measure constructed by the Wigner-Ville spectrum, is a unique measure for multivariate locally stationary processes.

3. Statistical Inference for Local Granger Causality

In this section, we develop the foundations for statistical inference for local Granger causality. The proofs of the theoretical results are relegated to Section S3.

3.1 Local Whittle estimation

Let $\{\boldsymbol{X}_{t,T}\}$ be the multivariate locally stationary process defined by (2.2) with the time-varying spectral density $\boldsymbol{f}(u,\lambda)$ defined by (2.4). The starting point is local estimation by fitting a parametric spectral density model $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$, to $\boldsymbol{f}(u,\lambda)$.

Consider the observation stretch $(\boldsymbol{X}_{1,T}, \dots, \boldsymbol{X}_{T,T})$ and define $\boldsymbol{I}_T(u,\lambda)$ to be the pre-periodogram matrix

$$\boldsymbol{I}_T(u,\lambda) = \frac{1}{2\pi} \sum_{\ell : 1 \le [uT+1/2 \pm \ell/2] \le T} \boldsymbol{X}_{[uT+1/2+\ell/2],T} \boldsymbol{X}^{\top}_{[uT+1/2-\ell/2],T} \exp(-\mathrm{i}\lambda\ell). \tag{3.16}$$

Note that the pre-periodogram $\boldsymbol{I}_T(u, \lambda)$ was first introduced in Neumann and von Sachs (1997).

We define the spectral divergence $\mathcal{L}(\boldsymbol{\theta}, u)$ between the parametric spectral density and the time-varying spectral density as

$$\mathcal{L}(\boldsymbol{\theta}, u) = \int_{-\pi}^{\pi} \ln \det \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda) + \operatorname{Tr}\Big(\boldsymbol{f}(u, \lambda) \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda)\Big) \,\mathrm{d}\lambda. \tag{3.17}$$

For any fixed $u \in [0, 1]$, define $\boldsymbol{\theta}_0(u)$ as

$$\boldsymbol{\theta}_0(u) := \arg\min_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta}, u). \tag{3.18}$$

Let $u_k := k/T$. The sample analogue $\mathcal{L}_T$ of the spectral divergence is defined as

$$\mathcal{L}_T(\boldsymbol{\theta}, u) = \frac{1}{T} \sum_{k=1}^{T} \frac{1}{b_T} K\Big(\frac{u - u_k}{b_T}\Big) \int_{-\pi}^{\pi} \ln \det \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda) + \operatorname{Tr}\Big(\boldsymbol{I}_T(u_k, \lambda) \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda)\Big) \,\mathrm{d}\lambda, \tag{3.19}$$

and the local Whittle estimator of $\hat{\boldsymbol{\theta}}_T(u)$ is defined as

$$\hat{\boldsymbol{\theta}}_T(u) := \arg\min_{\boldsymbol{\theta} \in \Theta} \mathcal{L}_T(\boldsymbol{\theta}, u). \tag{3.20}$$

Remark 3. In the literature, the local Whittle estimation has been applied not only to short-memory locally stationary processes but also to long-memory processes (e.g. Chan and Palma (2020)). Especially, Chan and Palma (2020) shares similarities with our approach in that they also established the central limit theorem by considering the higher-order cumulants of the local Whittle estimator. In contrast, we here consider the local Whittle estimator, incorporating a kernel function K, for multivariate locally stationary processes, whereas their work considered the estimator for univariate locally stationary long-memory processes by segmenting observations into blocks of size N

and bounding the coefficients by their shifts. To keep brevity of our paper, we list assumptions and the central limit theorem in the following, and relegate the proof to the Supplementary Materials.

We impose the following assumptions on the class of time-varying spectral densities and the kernel function K in (3.19) to investigate the asymptotic properties of the local Whittle estimator (3.20). As a remark, we may make the same assumptions about the mapping $u \mapsto A(u, j)$ since the smoothness condition of the time-varying spectral density matrix $\boldsymbol{f}(u, \lambda)$ directly follows from the regularity properties of $u \mapsto A(u, j)$.

Assumption 2.

(i) *The time-varying spectral density matrix $\boldsymbol{f}(u, \lambda)$ is continuously differentiable with respect to u for $u \in (0, 1)$.*

(ii) *$K : \mathbb{R} \to \mathbb{R}$ is a nonnegative, bounded symmetric continuous function of bounded variation with a compact support $[-1, 1]$ satisfying $\int K(x)\, \mathrm{d}x = 1$. Let*

$$K_b(x) := \frac{1}{b} K\Big(\frac{x}{b}\Big),$$

where $b := b_T \to 0$, as $T \to \infty$.

We now specify the regularity conditions for the parametric model $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$ and the local parameter $\boldsymbol{\theta}(u)$. For the brevity, let $\boldsymbol{\theta} := \boldsymbol{\theta}(u)$ when u does not matter.

Assumption 3.

(i) *For any fixed $u \in [0, 1]$, $\boldsymbol{\theta}(u) \in \Theta$, where Θ is a compact subset of $\mathbb{R}^d$.*

(ii) *For any fixed* $u \in [0,1]$, $\boldsymbol{f}_{\boldsymbol{\theta}^{(1)}(u)} \neq \boldsymbol{f}_{\boldsymbol{\theta}^{(2)}(u)}$ *on a set of positive Lebesgue measure, if* $\boldsymbol{\theta}^{(1)}(u) \neq \boldsymbol{\theta}^{(2)}(u)$.

(iii) *The parametric spectral density matrix* $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$ *is bounded away from 0 for each component, and is continuously differentiable with respect to* λ *for* $\lambda \in (-\pi, \pi)$.

(iv) *For any* $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{f}_{\boldsymbol{\theta}}$ *is positive definite and it is twice continuously differentiable with respect to* $\boldsymbol{\theta}$.

(v) *For any fixed* $u \in [0,1]$,

(v-a) $\boldsymbol{\theta}_0(u) \in \Theta$ *is the unique minimizer of* $\mathcal{L}(\boldsymbol{\theta}, u)$ *and lies in the interior of* Θ.

(v-b) *the following matrix is positive definite:*

$$M_f^u = \int_{-\pi}^{\pi} \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathrm{Tr}\Big\{ \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \boldsymbol{f}(u, \lambda) \Big\} + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln \det \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda) \right] \mathrm{d}\lambda$$

First, let us consider the asymptotics for the sample analog of the spectral divergence $\mathcal{L}_T(\boldsymbol{\theta}, u)$.

Theorem 1. *Suppose Assumptions 1, 2 and 3 hold. For any* $u \in (0,1)$, *if* $b_T^{-1} = o\big(T (\ln T)^{-6}\big)$ *and* $b_T = o(T^{-1/5})$, *then we have*

$$\sqrt{T b_T}\big(\mathcal{L}_T(\boldsymbol{\theta}, u) - \mathcal{L}(\boldsymbol{\theta}, u) \big) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}^{\mathcal{L}}(u)).$$

as $T \to \infty$, *where*

$$\mathbb{V}^{\mathcal{L}}(u) = 4\pi \int_{-1}^{1} K(v)^2 \,\mathrm{d}v \Bigg(\int_{-\pi}^{\pi} \mathrm{Tr}\Big(\boldsymbol{f}(u,\lambda) \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \boldsymbol{f}(u,\lambda) \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \Big) \,\mathrm{d}\lambda + \frac{1}{2} \sum_{r,t,v,w=1}^{p} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Big(\boldsymbol{f}_{\boldsymbol{\theta}}^{rt}(\lambda_1) \boldsymbol{f}_{\boldsymbol{\theta}}^{vw}(\lambda_2) \tilde{\gamma}_{rtvw}(u; -\lambda_1, \lambda_2, -\lambda_2) \Big) \,\mathrm{d}\lambda_1 \,\mathrm{d}\lambda_2 \Bigg), \quad (3.21)$$

where $\tilde{\gamma}$ is the fourth-order spectral density of the process.

Remark 4. Inspection of the proof of Theorem 1 reveals that the main order of bias is $O(T^{1/2}b_T^{5/2})$ and the asymptotic variance is of order $O(T^{-1}b_T^{-1})$. The optimal order of the bandwidth b_T can be determined by equating squared bias and variance. Thus, we obtain the optimal order $b_T = O(T^{-1/3})$ and the mean square error is $O(T^{-2/3})$. This optimal order is similar to the one derived in Künsch (1989) in the context of statistical inference for stationary time series.

Let $M^u_{f,0}$ be

$$M^u_{f,0} = \int_{-\pi}^{\pi} \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathrm{Tr}\Big\{ \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \boldsymbol{f}(u,\lambda) \Big\} + \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln\det \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} \mathrm{d}\lambda.$$

Now we establish the asymptotic normality of the local estimator $\hat{\boldsymbol{\theta}}_T(u)$.

Theorem 2. *Suppose Assumptions 1, 2 and 3 hold. For any $u \in (0,1)$, if $b_T^{-1} = o\big(T(\ln T)^{-6}\big)$ and $b_T = o(T^{-1/5})$, then we have*

$$\sqrt{Tb_T}\big(\hat{\boldsymbol{\theta}}_T(u) - \boldsymbol{\theta}_0(u)\big) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{V}(u)), \tag{3.22}$$

as $T \to \infty$, where $\mathbb{V}(u) := (M^u_{f,0})^{-1}\mathbb{V}^{\boldsymbol{\theta}}(u)(M^u_{f,0})^{-1}$ and

$$\begin{aligned}\mathbb{V}^{\boldsymbol{\theta}}(u)_{ab} = 4\pi &\int_{-1}^{1} K(v)^2\,\mathrm{d}v \\ &\times \left(\int_{-\pi}^{\pi} \mathrm{Tr}\Big[\boldsymbol{f}(u,\lambda) \Big\{ \frac{\partial}{\partial\theta_a} \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \Big\} \boldsymbol{f}(u,\lambda) \Big\{ \frac{\partial}{\partial\theta_b} \boldsymbol{f}_{\boldsymbol{\theta}}^{-1}(\lambda) \Big\} \Big]\,\mathrm{d}\lambda \right. \\ + \frac{1}{2} \sum_{r,t,v,w=1}^{p} &\left. \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \Big(\frac{\partial}{\partial\theta_a} \boldsymbol{f}_{\boldsymbol{\theta}}^{rt}(\lambda_1) \cdot \frac{\partial}{\partial\theta_b} \boldsymbol{f}_{\boldsymbol{\theta}}^{vw}(\lambda_2) \tilde{\gamma}_{rtvw}(u; -\lambda_1, \lambda_2, -\lambda_2) \Big)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2 \right),\end{aligned} \tag{3.23}$$

$(a, b = 1, \ldots, d)$, where $\tilde{\gamma}$ is the fourth-order spectral density of the process.

3.2 Inference for causality measures

In this subsection, we develop the statistical inference for the local Granger causality measure (2.6) based on the parametric model $\{\boldsymbol{f}_{\boldsymbol{\theta}(u)} \mid \boldsymbol{\theta}(u) \in \Theta\}$. Denote by $\Sigma_{\boldsymbol{\theta}(u)}$ the parametric one-step-ahead prediction error matrix, and by $\boldsymbol{g}_{\boldsymbol{\theta}(u)}$ the parametric model for the companion process (2.5). Note that the matrix $\boldsymbol{g}_{\boldsymbol{\theta}(u)}$ is uniquely determined by the model $\boldsymbol{f}_{\boldsymbol{\theta}(u)}$ and the matrix $\Sigma_{\boldsymbol{\theta}(u)}$ (see, e.g., (2.7) and (2.10)).

Suppose $\boldsymbol{f}_{\boldsymbol{\theta}(u)}$, $\Sigma_{\boldsymbol{\theta}(u)}$ and $\boldsymbol{g}_{\boldsymbol{\theta}(u)}$ have the same partition as (2.4). To make the statistical inference feasible, we impose the following assumption on the parametric models.

Assumption 4. *For any* $\boldsymbol{\theta} \in \Theta$,

$$\int_{-\pi}^{\pi} \ln\lvert \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)\rvert \, \mathrm{d}\lambda > -\infty.$$

Assumption 4 is usually referred to as the maximal rank condition. This condition guarantees that the fitted model is full-rank and the model does not contain a perfectly predictable process; Otherwise, the Granger causality measure (2.6) for the fitted model may be not well-defined for the case that the denominator of the fraction inside the logarithm is 0. Under Assumption 4, $\boldsymbol{f}_{\boldsymbol{\theta}(u)}(\lambda)$ is non-degenerate a.e. and $\Sigma_{\boldsymbol{\theta}(u)}$ is positive definite for any fixed $u \in [0, 1]$. Now the parametric local Granger causality for (2.6) is

$$\mathrm{GC}^{(2\to 1)}(u;\, \boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{FGC}\big(\lambda; \boldsymbol{\theta}(u)\big) \, \mathrm{d}\lambda, \tag{3.24}$$

where

$$\mathrm{FGC}(\lambda;\, \boldsymbol{\theta}) = \ln \frac{\lvert \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11}\rvert}{\left\lvert \boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11} - 2\pi \boldsymbol{g}_{\boldsymbol{\theta}}(\lambda)_{12} \tilde{\Sigma}_{\boldsymbol{\theta},22}^{-1} \boldsymbol{g}_{\boldsymbol{\theta}}(\lambda)_{21} \right\rvert}.$$

The main results are described in the following.

Theorem 3. *Suppose Assumptions 1, 2, 3 and 4 hold. If we have, for some* $i = 1, \ldots, d$,

$$\frac{\partial}{\partial \theta_i} \mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to 1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} \neq 0, \qquad \text{for some } \lambda \in (-\pi, \pi], \tag{3.25}$$

uniformly in $u \in [0,1]$, *and if* $b_T^{-1} = o\big(T(\ln T)^{-6}\big)$ *and* $b_T = o(T^{-1/5})$, *then we have*

$$\sqrt{Tb_T}\Big(\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T) - \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0)\Big) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}^{GC}(u)),$$

where

$$\mathbb{V}^{GC}(u) = \Big(\nabla \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0)\Big)^\top \mathbb{V}(u) \left(\nabla \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0)\right),$$

and

$$\nabla \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0) = \left(\frac{\partial}{\partial \theta_1} \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0), \ldots, \frac{\partial}{\partial \theta_d} \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0)\right)^\top$$

with

$$\frac{\partial}{\partial \theta_i} \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \mathrm{FGC}_{\boldsymbol{\theta}(u)}^{(2\to 1)}(\lambda)\, \mathrm{d}\lambda.$$

There are situations when condition (3.25) may not be satisfied. That is, for some $u \in (0,1)$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to 1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} = \mathbf{0}, \qquad \text{a.e. } \lambda \in (-\pi, \pi]. \tag{3.26}$$

In this case, we centralize $\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T)$ as $\mathrm{CGC}(u;\, \hat{\boldsymbol{\theta}}_T)$, i.e.,

$$\mathrm{CGC}(u;\, \hat{\boldsymbol{\theta}}_T) := \mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T) - \mathrm{GC}^{(2\to 1)}(u; \boldsymbol{\theta}_0). \tag{3.27}$$

Then we have the following result.

Theorem 4. *Suppose that Assumptions 1, 2, 3 and 4 hold. In addition, assume* (3.26) *with*

$$\mathcal{H}(u,\lambda) := \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} \neq O_{d\times d}, \qquad \text{for some } \lambda\in(-\pi,\pi],$$

for $u\in(0,1)$. Let

$$\mathcal{H}(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{H}(u,\lambda)\,\mathrm{d}\lambda. \tag{3.28}$$

Then if $b_T^{-1} = o\big(T(\ln T)^{-6}\big)$ and $b_T = o(T^{-1/5})$, the following result holds

$$2Tb_T\mathrm{CGC}^{(2\to1)}(u;\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \mathcal{N}\big(0,\mathbb{V}(u)\big)^\top\mathcal{H}(u)\mathcal{N}\big(0,\mathbb{V}(u)\big), \tag{3.29}$$

where the normal distribution $\mathcal{N}\big(0,\mathbb{V}(u)\big)$ is defined in Theorem 2. In particular, if $\mathbb{V}(u)^{-1/2}\mathcal{H}(u)\mathbb{V}(u)^{-1/2}$ is an idempotent matrix, then the right hand side of (3.29) *has a chi-squared distribution χ^2_ν with the degrees of freedom*

$$\nu = \mathrm{Tr}(\mathbb{V}(u)^{-1/2}\mathcal{H}(u)\mathbb{V}(u)^{-1/2}).$$

Let us summarize Assumptions 1–4 before we give an example satisfying these assumptions. Assumption 1 is a condition for innovation processes. Assumptions 2–3 are regularity conditions for the local Whittle estimation. They could be replaced by other regularity conditions (e.g. regularity conditions in Chan and Palma (2020)) for the parameter estimation by the local Whittle estimator. Assumption 4 is a condition to ensure that the fitted model is full-rank and the model is not a perfectly predictable process by a linear operation on the past. See Hannan (1970, Chapter III, Section 5) for perfectly predictable processes.

Example 1 (Time-varying vector autoregression model)**.** Suppose the multivariate locally Gaussian stationary process (2.2) has the time-varying spectral density

$$\boldsymbol{f}(u,\lambda) = \frac{1}{2\pi}\big(I_2 + A(u)\exp(\mathrm{i}\lambda)\big)^{-1}\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\big(I_2 + A(u)^\top \exp(-\mathrm{i}\lambda)\big)^{-1},$$

where $A(u) = \begin{pmatrix}\alpha_{11}(u) & \alpha_{12}(u)\\ \alpha_{21}(u) & \alpha_{22}(u)\end{pmatrix}$ and $\alpha_{12}(u) \equiv 0$.

We adopt the following parametric spectral density $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$ for model fitting.

$$\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda) = \frac{1}{2\pi}\left(I_2 + \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\exp(\mathrm{i}\lambda)\right)^{-1} \times \begin{pmatrix}s_{11} & s_{12}\\ s_{12} & s_{22}\end{pmatrix}\left(I_2 + \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}^\top \exp(-\mathrm{i}\lambda)\right)^{-1},$$

where $\boldsymbol{\theta} = (a_{11}, a_{12}, a_{21}, a_{22}, s_{11}, s_{12}, s_{22})^\top$. From the definition (3.24), we have

$$\mathrm{GC}^{(2\to1)}(u;\,\boldsymbol{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} -\ln\left|1 - 2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21} f_{\boldsymbol{\theta}}(\lambda)_{11}^{-1}\right| \mathrm{d}\lambda.$$

Thus, $\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)$ is

$$\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda) = -\ln\left|1 - 2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21} f_{\boldsymbol{\theta}}(\lambda)_{11}^{-1}\right|.$$

A straightforward computation leads to

$$\frac{\partial}{\partial\boldsymbol{\theta}}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)\Big|_{a_{12}=0} = \mathbf{0}, \qquad \text{for any } \lambda \in (-\pi,\pi].$$

In addition, it holds that

$$\frac{\partial^2}{\partial a_{12}^2}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} = \frac{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2}{\sigma_{11}^4}\frac{|1 - \alpha_{11}(u)\exp(\mathrm{i}\lambda)|^2}{|1 - \alpha_{22}(u)\exp(\mathrm{i}\lambda)|^2} > 0;$$

and for all $\lambda \in (-\pi, \pi]$,

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)} = 0, \quad \text{for } \theta_i \neq a_{12} \text{ or } \theta_j \neq a_{12}.$$

Let $\hat{\boldsymbol{\theta}}_T = \big(\hat{\alpha}_{11}(u), \hat{\alpha}_{12}(u), \hat{\alpha}_{21}(u), \hat{\alpha}_{22}(u), \hat{\sigma}_{11}, \hat{\sigma}_{12}, \hat{\sigma}_{22}\big)^\top$ be the local Whittle estimator defined in (3.20). Applying Theorem 4, we obtain

$$Tb_T \frac{\sigma_{11}^4\Big(\int_{-1}^1 K(v)^2\,\mathrm{d}v\Big)^{-1}}{\big(1+\alpha_{11}(u)^2 - 2\alpha_{11}(u)\alpha_{22}(u)\big)\big(\sigma_{11}\sigma_{22}-\sigma_{12}^2\big)^2}\mathrm{CGC}(u;\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi_1^2, \tag{3.30}$$

since

$$\begin{aligned}\mathcal{H}(u)_{22} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial a_{12}^2}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to1)}(\lambda)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0(u)}\mathrm{d}\lambda \\ &= \frac{2\big(1+\alpha_{11}(u)^2-2\alpha_{11}(u)\alpha_{22}(u)\big)\big(\sigma_{11}\sigma_{22}-\sigma_{12}^2\big)^2}{\big(1-\alpha_{22}(u)^2\big)\sigma_{11}^4},\end{aligned}$$

and

$$\mathbb{V}(u)_{22} = \big(1-\alpha_{22}(u)^2\big)\int_{-1}^1 K(v)^2\,\mathrm{d}v.$$

3.3 Hypothesis testing for causality measures

We now address the hypothesis testing problem for the local measures $\mathrm{GC}^{(2\to1)}$. Suppose that we want to test for local causality at a particular rescaled time $u \in [0,1]$. Define the local hypothesis $H_0^{(2\to1)}$ to be

$$H_0^{(2\to1)} : \mathrm{GC}^{(2\to1)}(u) = c. \tag{3.31}$$

We consider two cases of the null hypothesis (3.31): (i) $c = 0$, and (ii) $c > 0$.

Let us first consider the case (i) $c = 0$. For any fixed $u \in [0,1]$, with the shorthand $\boldsymbol{\theta} = \boldsymbol{\theta}(u)$, we have

$$\mathrm{GC}^{(2\to1)}(u;\, \boldsymbol{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \frac{|\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11}|}{\left|\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11} - 2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21}\right|}\, \mathrm{d}\lambda,$$

and thus, $\mathrm{GC}^{(2\to1)}(u;\, \boldsymbol{\theta}) = 0$ if and only if

$$|\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11}| = \left|\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11} - 2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21}\right|. \tag{3.32}$$

Here, $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)_{11}$ and $2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21}$ are both Hermitian. Thus, the equality (3.32) holds if

$$2\pi g_{\boldsymbol{\theta}}(\lambda)_{12}\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1} g_{\boldsymbol{\theta}}(\lambda)_{21} = O_{m\times m}.$$

It follows that $g_{\boldsymbol{\theta}}(\lambda)_{12} = O_{m\times M}$, since $\left(\tilde{\Sigma}_{\boldsymbol{\theta},22}\right)^{-1}$ is positive definite. Since $g_{\boldsymbol{\theta}}(\lambda)_{12} = O_{m\times M}$, it is straightforward to see that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{FGC}^{(2\to1)}_{\boldsymbol{\theta}}(\lambda) = \mathbf{0},$$

which is the case we considered in Theorem 4.

Accordingly, for the local hypothesis $H_0^{(2\to1)} : \mathrm{GC}^{(2\to1)}(u) = 0$, we take

$$S^{\dagger}(u) := 2T\, b_T\, \mathrm{GC}^{(2\to1)}(u;\, \hat{\boldsymbol{\theta}}_T)$$

as the test statistic. Then we have the following result.

Theorem 5. *Suppose Assumptions 1, 2, 3 and 4 hold. Under the null hypothesis* $H_0^{(2\to1)} : \mathrm{GC}^{(2\to1)}(u) = 0$, *if* $b_T^{-1} = o\big(T(\ln T)^{-6}\big)$ *and* $b_T = o(T^{-1/5})$, *it holds that*

$$S^{\dagger}(u) \xrightarrow{d} \mathcal{N}\big(0, \mathbb{V}(u)\big)^{\top} \mathcal{H}(u) \mathcal{N}\big(0, \mathbb{V}(u)\big).$$

where $\mathcal{N}\big(0, \mathbb{V}(u)\big)$ *is defined in Theorem 2.*

Remark 5. The matrix $\mathcal{H}(u)$ in (3.28) is unknown in general, but it is determinable from the parameterization of $\boldsymbol{f}_{\boldsymbol{\theta}}(\lambda)$. In practice, the matrix $\mathcal{H}(u)$ should be replaced with its plug-in version

$$\mathcal{H}(u;\, \hat{\boldsymbol{\theta}}_T) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\mathrm{FGC}_{\boldsymbol{\theta}}^{(2\to 1)}(\lambda)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_T(u)}\,\mathrm{d}\lambda$$

to construct an asymptotic $(1-\alpha)$ confidence interval for $\mathrm{GC}^{(2\to 1)}(u)$. Instead, in some situation, for instance, as a continuation of Example 1, if we take

$$\tilde{S}^\dagger(u) := T\, b_T\, \frac{\hat{\sigma}_{11}^4\left(\int_{-1}^1 K(v)^2\,\mathrm{d}v\right)^{-1}}{\left(1+\hat{\alpha}_{11}(u)^2 - 2\hat{\alpha}_{11}(u)\hat{\alpha}_{22}(u)\right)\left(\hat{\sigma}_{11}\hat{\sigma}_{22}-\hat{\sigma}_{12}^2\right)^2}\mathrm{CGC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T)$$

as a sample version of the left hand side in (3.30), then the confidence interval is $[0, \chi^2_{1,1-\alpha}]$, where $\chi^2_{1,1-\alpha}$ denotes the $(1-\alpha)$ quantile of the chi-squared distribution with 1 degree of freedom. It is straightforward to see that the test by $\tilde{S}^\dagger(u)$ is consistent. If the null hypothesis is rejected, then the conclusion is that at level α there is sufficient evidence to suggest that there exists local Granger causality from one series to another at rescaled time $u \in [0, 1]$.

We now move on to the second case (ii) $c > 0$. The Wald type test statistic $S^*(u)$ is

$$\begin{aligned} &S^*(u) = Tb_T\left(\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T) - c\right) \\ &\times\left[\left(\nabla\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T)\right)\mathbb{V}(u)\left(\nabla\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T)\right)^\top\right]^{-1}\left(\mathrm{GC}^{(2\to 1)}(u;\, \hat{\boldsymbol{\theta}}_T) - c\right). \end{aligned} \tag{3.33}$$

The following result is a direct consequence of Theorem 3.

Theorem 6. *Suppose Assumptions 1, 2, 3 and 4 hold. Under the null hypothesis* $H_0^{(2\to1)}: \mathrm{GC}^{(2\to1)}(u) = c > 0$, *if* $b_T^{-1} = o\big(T(\ln T)^{-6}\big)$ *and* $b_T = o(T^{-1/5})$, *we have*

$$S^*(u) \xrightarrow{d} \chi_d^2,$$

where χ_d^2 *is a chi-squared distribution with the degrees of freedom* d*.*

Remark 6. The covariance matrix $\mathbb{V}(u)$ in (3.33) is usually unknown, and thus, we have to construct a consistent estimator $\hat{\mathbb{V}}(u)$ instead. If the process $\{\boldsymbol{X}_{t,T}\}$ is non-Gaussian, we have to take the estimation of the fourth-order spectral density of the process into account. We refer to Taniguchi (1982) or Keenan (1987) for further details.

4. Data Analysis

In this section, we apply local Granger causality to two real datasets – EEG data and financial data.

4.1 EEG data

We provide a brief description of the data. The EEG signals are sampled at the rate of 100 Hertz. The recordings are taken from channels the central channels (C3, C4, Cz), parietal channels (P3, P4) and the temporal channels (T3, T4, T5) which correspond roughly to the central, parietal and temporal brain cortical regions (See Figure 1). The original dataset has 32680 time points for each channel (i.e., the period of the observation is 326.8 seconds). This dataset was previously analyzed in Ombao et al. (2005) and Schröder and Ombao (2019). However, none of these two papers addressed the very important issue of causality. This is the first paper that examined local Granger causality features in this data.

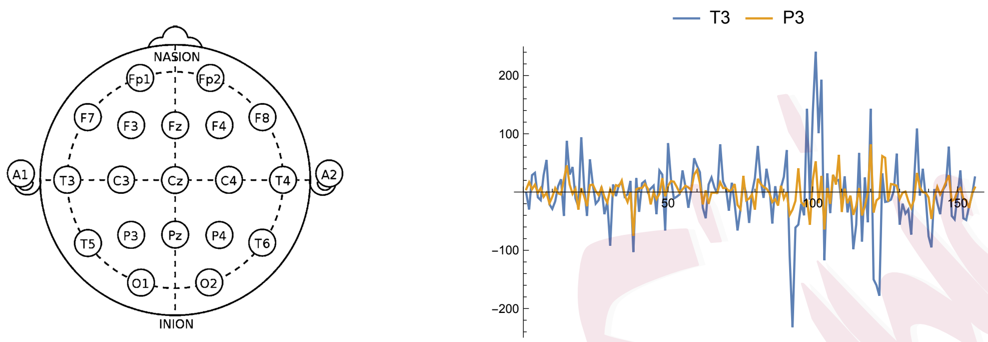

Figure 1: EEG channels (left) and the data of channels T_3 and P_3 (right).

Local Granger causality was estimated and tested at every rescaled time point $u_k = 2.1k/326.8$ and $u_k = 4.2k/326.8$, due to the computational cost of the local Whittle estimation. This is equivalent to estimating and testing every 2.1 and 4.2 seconds respectively. We refer to these partial data as one at regular intervals of 2.1 seconds and 4.2 seconds. As for the Whittle estimation, we locally fitted VARMA(1, 1) model to the data and took the bandwidth b_T as $b_T = T^{-1/4}$, where T is the length of theses partial data. In Figure 2, we show the logarithm of local Granger causality between two specific channels P3 (left parietal) and T3 (left temporal). These two channels are of primary interest because the patient suffered from left temporal lobe epilepsy - though the precise location is quite close to the parietal lobe. Thus the seizure focus is the left temporal lobe and any abnormalities in the EEG are captured in the T3 and P3 channels. The 95% confidence intervals are shown below: the dashed one is computed from the data at regular intervals of 4.2 seconds; the dotted one is computed from the data at regular intervals of 2.1 seconds.

The partial data at regular intervals of 2.1 seconds and 4.2 seconds share a very similar move of the logarithm of local Granger causality and the 95% confidence intervals are also very similar. In general, the higher temporal resolution the sampling is, theoretically more accurate the estimates are. In our simulation results, the estimates from the data at regular intervals of 2.1 seconds are as good as that of 4.2 seconds. The analysis suggests that T3 does not cause P3 in the Granger sense, but P3 causes T3 in the Granger sense at latest after the rescaled time $u = 0.15$. This is a quite interesting finding because previous analyses have focused on the T3 channel because of the distinctly large amplitudes immediately post-seizure onset. However, the novel finding here is that the direction of Granger causality actually flows from P3 to T3. This suggests that, despite the relatively lower amplitude changes in P3, it still explains the future large amplitude fluctuations in T3.

Next, we further investigate the local causalities between the central channels C3, Cz and C4 at regular intervals of 4.2 seconds. Figure 3 represents the numerical results of the causality from the column to the row. For example, the middle plot in the first row shows the causality from the channel C3 to Cz. Still, the 95% confidence intervals are below the dashed red lines and the logarithms of local Granger causalities are shown by the dashed black lines. From Figure 3, the conclusion is that channel the left central channel C3 does not cause central channel Cz. Moreover, the right central channel C4 does not cause channels C3 and Cz - in the local Granger sense. In other cases, local Granger causality changes across the evolution of the epileptic seizure which confirms the dynamic activity of the brain. This is a new finding since all previous analyses were limited to modeling dependence using only coherence which accounts for contemporaneous dependence; that is, there was no phase or lead-lag analysis. Moreover, this novel finding is quite interesting because a change in the causality structure was captured even before the onset of the epileptic seizure, which was approximately at $u = 0.5$.

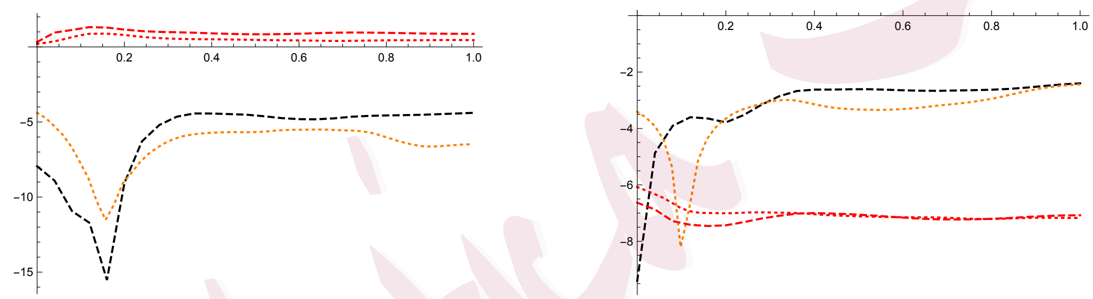

Figure 2: The logarithm of local Granger causality (LLGC) from the channel T3 to P3 (left) and that from the channel P3 to T3 (right) during the rescaled time $u \in [0, 1]$. The LLGC for data at regular intervals of 4.2 seconds is shown in black and that for those of 2.1 seconds are shown in orange. The dashed red line shows the 95% confidence levels computed from the data at regular intervals of 4.2 seconds, while the dotted red line shows the 95% confidence levels computed from the data at regular intervals of 2.1 seconds.

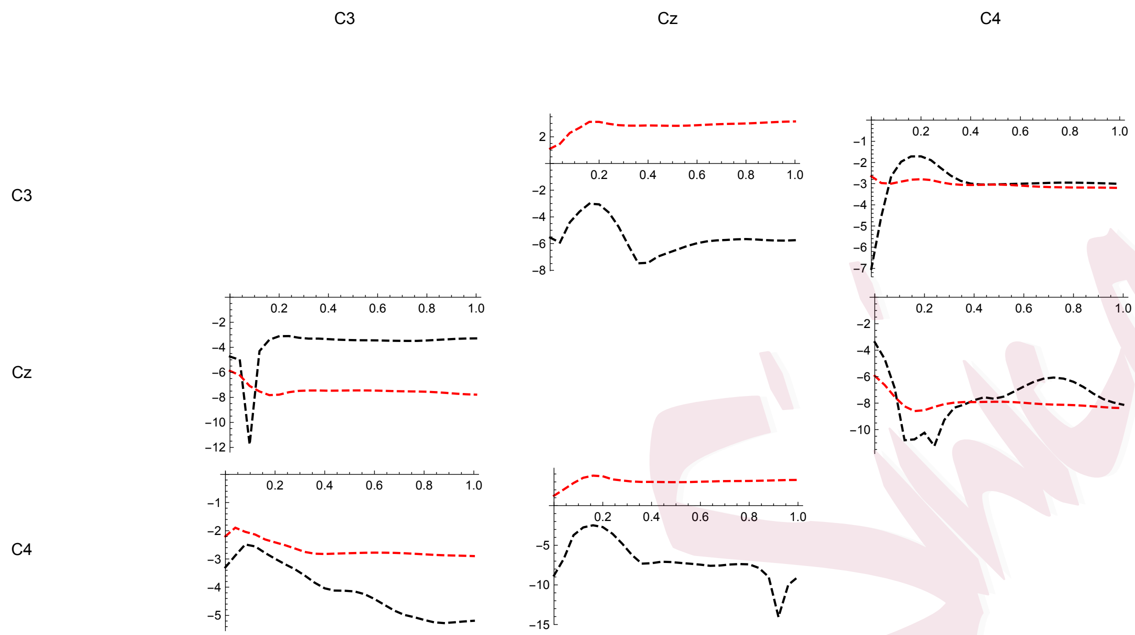

Figure 3: Plots of the logarithms of local Granger causality from the channels C3, Cz, C4 in the column to the channels C3, Cz, C4 in the row. The 95% confidence intervals are below the dashed red lines; The dashed black lines show the logarithms of local Granger causalities.

Similarly, we studied the causalities between the temporal channels T3, T5, T4 at regular intervals of 4.2 seconds. The plots of the numerical results are shown in Figure 4. Remember that the channels T3 and T4 are symmetrically located at both the left and right temporal cortical regions, respectively. The plots suggest that T3 and T4 do not cause each other in the local Granger sense. Furthermore, the channel T5, also on the left temporal cortical region, uniformly causes T3 in the data which is another interesting novel finding. It is already known to the neurologist that the patient has left temporal lobe epilepsy and that seizure events are generally initiated in the ”left temporal”l region (which is the area covered by the T3 and T5 channels). Using the novel proposed concept of local Granger causality, the analysis produced a highly specific result of brain functional connectivity, that is, the direction goes from T5 to T3 and not the other way around. As an additional result, P3 and P4 do not cause each other uniformly in the local Granger sense at 95% confidence level.

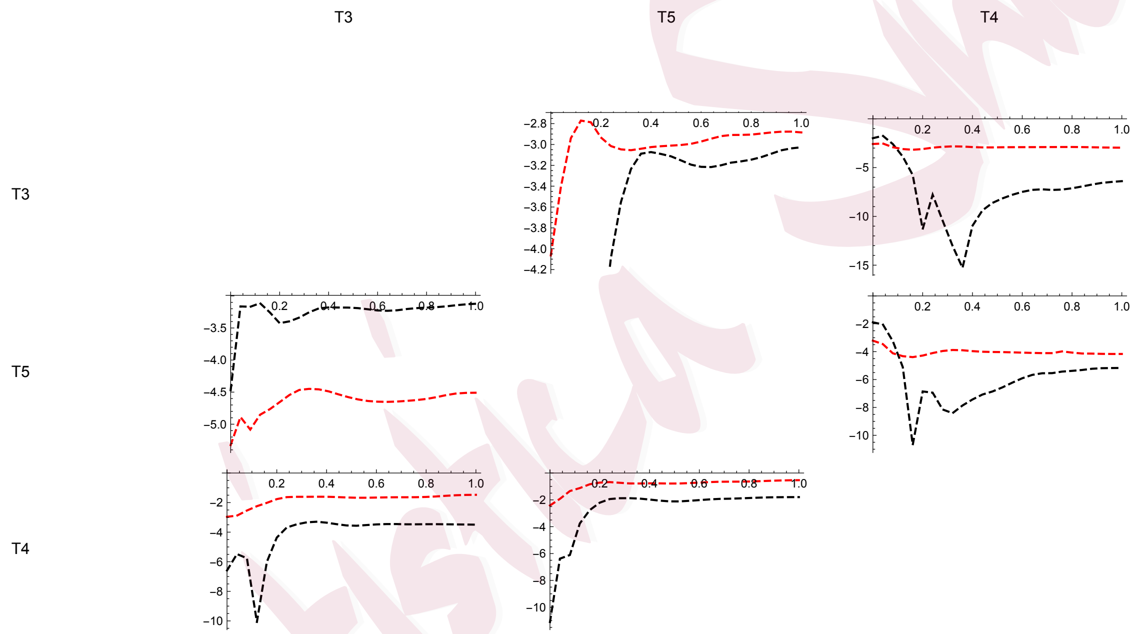

Figure 4: Plots of the logarithms of local Granger causality from the channels T3, T5, T4 in the column to the channels T3, T5, T4 in the row. The 95% confidence intervals are below the dashed red lines; The dashed black lines show the logarithms of local Granger causalities.

4.2 Stock market data

The dataset is the weekly log-returns of the closing stock prices of two financial groups (Mitsubishi and Mizuho) in the Nikkei index. For brevity, two financial groups are denoted by A and B here. The weekly data are from 2006 January 1st to 2010 December 26th, so the length of the data is $T = 260$. In this data analysis, we fitted VARMA(2, 0) model locally to the data. Similar to the bandwidth choice in EEG data analysis, we also used the bandwidth b_T as $b_T = T^{-1/4}$ for these stock market data.

In our data analysis, we compute local Granger causality $\mathrm{GC}^{\mathrm{A}\to\mathrm{B}}(u)$ and $\mathrm{GC}^{\mathrm{B}\to\mathrm{A}}(u)$ for rescaled time $u \in \{1/T, \ldots, 1\}$. In general, Granger causality is not symmetric (e.g., the analysis of EEG data) and we regard

$$\left\{ \left(\mathrm{GC}^{\mathrm{A}\to\mathrm{B}}(u), \mathrm{GC}^{\mathrm{B}\to\mathrm{A}}(u) \right) \in \mathbb{R}^2 \right\}_{u=\frac{1}{T},\ldots,1} \tag{4.34}$$

as a point cloud in $\mathbb{R}^2$. We separate local Granger causality in (4.34) into two parts: (1) from 2006 January 1st to 2008 June 29th; (2) from 2008 July 6th to 2010 December 26th. The part (1) and part (2) have the same length, i.e., the length of each part is 130.

We apply computational topology tools, *persistence diagram*, *persistence barcode*, *persistence landscape*, to capture the feature of these data points of local Granger causality (See, e.g. Fasy et al. (2014), for the details of the persistence diagram and persistence barcode). The plots of the persistence diagram and the persistence barcode are shown in Figure 5. The permutation test is applied to the point clouds (1) and (2) to test for equal topologies of the point clouds (1) and (2). In other words, the null hypothesis is no statistical difference between the persistence landscapes of local Granger causality. The result is statistically significant at the significance level of 0.01. It is known that there is a financial crisis between 2007 and 2008. Through the analysis of local Granger causality, we detected the structural change of the causality between these two financial data.

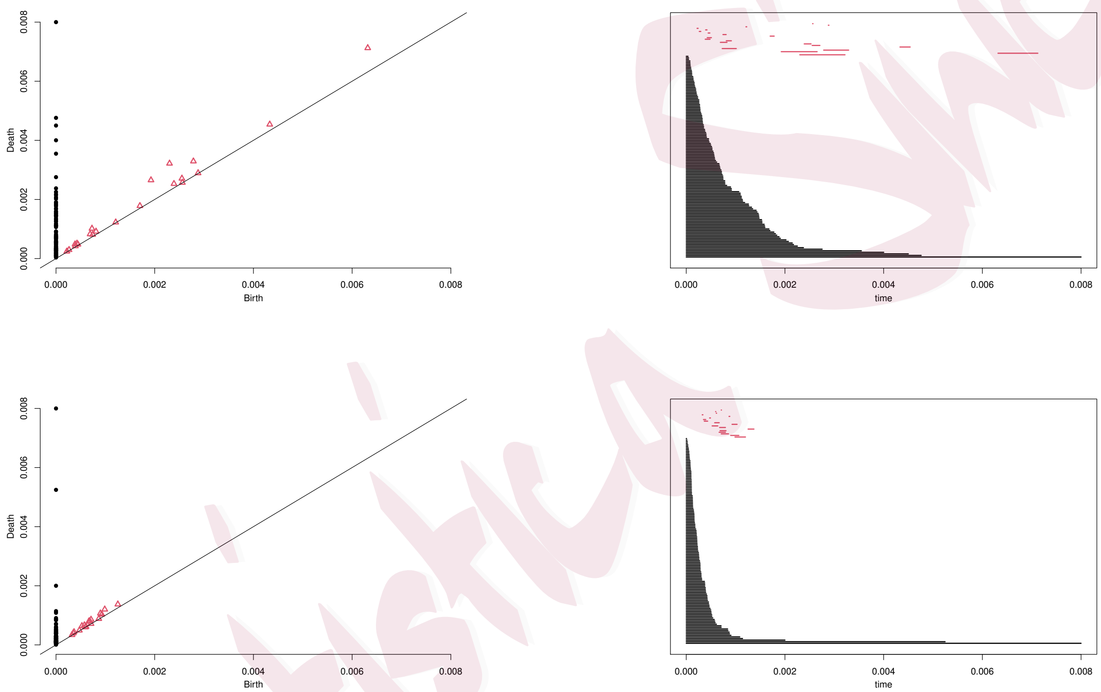

Figure 5: (Above) the persistence diagram of local Granger causality (left); and the persistence barcode of local Granger causality (right) during the period (1). (Below) the persistence diagram of local Granger causality (left); and the persistence barcode of local Granger causality (right) during the period (2).

5. Conclusion

The primary contribution of this paper is statistical inference for local Granger causality for multivariate time series under the framework of multivariate locally stationary processes. Our proposed concept of local Granger causality is a generalization of Geweke's measure and Hosoya's measure - both of which were developed only for stationary processes. Our proposed generalization is well characterized in the frequency domain and has the advantage of being able to capture time-evolving causality relationships. We developed a procedure for hypothesis testing for the existence of the local Granger causality from a parametric point of view. We demonstrate, through the analysis of real data, the efficiency and efficacy of this procedure to find the time-evolving aspects of the local Granger causality, which could be overlooked by the existing method for causality analysis. Our current approach to data analysis involves hypothesis testing on a coordinate-by-coordinate basis, presenting challenges associated with multiple testings. However, it is important to highlight that we proposed the Granger causality under a local time frame u. Testing at u implies causality within the local time frame, reflecting the characteristics of locally stationary processes. This makes the testing meaningful and significant.

In summary, we proposed a consistent method to detect the time change of the local Granger causality. While our proposed method is nonparametric, we note that a procedure for stationary processes was developed in Taniguchi et al. (1996). This could serve as an inspiration for constructing a nonparametric method for locally stationary processes to test for the local Granger causality, and compare the performance of both approaches. For multiple time series, to investigate the time change of the local Granger causality also suffers from the curse of dimensionality. There are many remaining challenges including dimension reduction in terms of causality between each component of multiple time series. In addition to the Lasso method in Tibshirani (1996), most penalized estimation procedures could be added to our parametric approach to shrink some minor causality between components. A frequency-specific local causality approach will also be elucidated in our future work.

Supplementary Materials

The online supplementary materials provide some technical parts of the paper due to the space limit.

Acknowledgements

The authors would like to thank the two anonymous referees for their constructive suggestions. This paper has benefited considerably from those comments. The first author gratefully acknowledge JSPS Grant-in-Aid for Young Scientists (B) 17K12652 and JSPS Grant-in-Aid for Scientific Research (C) 20K11719. The second author gratefully acknowledge JSPS Grant-in-Aid for Scientific Research (S) 18H05290 The third author gratefully acknowledge the KAUST Research Fund. The first two authors also would like to express their thanks to the Institute for Mathematical Science (IMS), Waseda

University, for their support for this research.

References

Chan, N. H. and W. Palma (2020). On the estimation of locally stationary long-memory processes. *Statistica Sinica 30*(1), 111–134.

Clarke, J. A. and S. Mirza (2006). A comparison of some common methods for detecting granger noncausality. *Journal of Statistical Computation and Simulation 76*(3), 207–231.

Dahlhaus, R. (1996). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications 62*(1), 139–168.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics 28*(6), 1762–1794.

Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics 151*(2), 101–112.

Dahlhaus, R. and L. Giraitis (1998). On the optimal segment length for parameter estimates for locally stationary time series. *Journal of Time Series Analysis 19*(6), 629–655.

Dahlhaus, R. and W. Polonik (2009). Empirical spectral processes for locally stationary time series. *Bernoulli 15*(1), 1–39.

Dolado, J. J. and H. Lütkepohl (1996). Making wald tests work for cointegrated var systems. *Econometric Reviews 15*(4), 369–386.

Fasy, B. T., F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh (2014). Confidence sets for persistence diagrams. *The Annals of Statistics 42*(6), 2301–2339.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association 77*(378), 304–313.

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association 79*(388), 907–915.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society 37*(3), 424–438.

Granger, C. W. and J.-L. Lin (1995). Causality in the long run. *Econometric Theory 11*(3), 530–536.

Granger, C. W. J. (1963). Economic processes involving feedback. *Information and Control 6*(1), 28–48.

Hannan, E. J. (1970). *Multiple Time Series*, Volume 38. John Wiley & Sons.

Hosoya, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary processes. *Probability Theory and Related Fields 88*(4), 429–444.

Huang, H.-Y., H. Ombao, and D. S. Stoffer (2004). Discrimination and classification of nonstationary time series using the slex model. *Journal of the American Statistical Association 99*(467), 763–774.

Keenan, D. M. (1987). Limiting behavior of functionals of higher-order sample cumulant spectra. *The Annals of Statistics 15*(1), 134–151.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics 17*(3), 1217–1241.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer Science & Business Media.

Martin, W. and P. Flandrin (1985). Wigner-ville spectral analysis of nonstationary processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing 33*(6), 1461–1470.

Neumann, M. H. and R. von Sachs (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics 25*(1), 38–76.

Ombao, H., R. von Sachs, and W. Guo (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association 100*(470), 519–531.

Ombao, H. C., J. A. Raz, R. von Sachs, and B. A. Malow (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association 96*(454), 543–560.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological) 27*(2), 204–229.

Rozanov, I. A. I. A. (1967). *Stationary random processes.* Holden-Day series in time series analysis. San Francisco: Holden-Day.

Sakiyama, K. and M. Taniguchi (2004). Discriminant analysis for locally stationary processes. *Journal of Multivariate Analysis 90*(2), 282–300.

Schröder, A. and H. Ombao (2019). Fresped: Frequency-specific change-point detection method in multi-channel epileptic seizure eeg data. *Journal of the American Statisical Association 114*, 115–128.

Shumway, R. H. and D. S. Stoffer (2000). *Time series analysis and its applications.* New York: Springer-Verlag.

Sims, C. A., J. H. Stock, and M. W. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica: Journal of the Econometric Society 58*(1), 113–144.

Taniguchi, M. (1982). On estimation of the integrals of the fourth order cumulant spectral density. *Biometrika 69*(1), 117–122.

Taniguchi, M. and Y. Kakizawa (2000). *Asymptotic theory of statistical inference for time series.* Springer Science & Business Media.

Taniguchi, M., M. L. Puri, and M. Kondo (1996). Nonparametric approach for non-gaussian vector stationary processes. *Journal of Multivariate Analysis 56*(2), 259–283.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B 58*(1), 267–288.

Toda, H. Y. and P. C. Phillips (1994). Vector autoregression and causality: a theoretical overview and simulation study. *Econometric Reviews 13*(2), 259–285.

Toda, H. Y. and P. C. B. Phillips (1993). Vector autoregressions and causality. *Econometrica 61*, 1367–1393.

Yamada, H. and H. Y. Toda (1998). Inference in possibly integrated vector autoregressive models: Some finite sample evidence. *Journal of Econometrics 86*(1), 55–95.

Yan Liu

Faculty of Science and Engineering, Waseda University

3-4-1, Okubo, Shinjuku-ku, Tokyo 169-8555, Japan.

E-mail: liu@waseda.jp

Masanobu Taniguchi

Faculty of Science and Engineering, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Email: taniguchi@waseda.jp

Hernando Ombao

Statistics Program, King Abdullah University of Science and Technology

Thuwal 23955-6900, Saudi Arabia.

Email: hernando.ombao@kaust.edu.sa